GenCore version 5.1.7

OM nucleic - nucleic search, using sw model

Run on: February 17, 2006, 09:49:14 ; Search time 1478 Seconds
(without alignments)
11277.670 Million cell updates/sec

Title: US-10-607-806-1-C7256_COPY_7000_9500
Perfect score: 2499
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_21:*
1: geneseqn1980s:*
2: geneseqn1990s:*
3: geneseqn2000s:*
4: geneseqn2001as:*
5: geneseqn2001bs:*
6: geneseqn2002as:*
7: geneseqn2002bs:*
8: geneseqn2003as:*
9: geneseqn2003bs:*
10: geneseqn2003cs:*
11: geneseqn2003ds:*
12: geneseqn2004as:*
13: geneseqn2004bs:*
14: geneseqn2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2498.6	100.0	12174	12	ADI35082	Adi35082 Human PLA
	2	2497.4	99.9	12174	12	ADJ09983	Adj09983 Human pho
	3	2497.4	99.9	13612	6	ABK47376	Abk47376 Human Pho
c	4	456.6	18.3	42360	13	ABD33466	Abd33466 Human can
	5	434.2	17.4	93544	13	ABD33504	Abd33504 Human can
c	6	434.2	17.4	160482	11	ACN43914	Acn43914 Human gen
c	7	433.2	17.3	91760	11	ACN44410	Acn44410 Human gen
c	8	425.4	17.0	38538	13	ABD33357	Abd33357 Human can
	9	417.2	16.7	122888	6	ABK83569	Abk83569 Human cDN
c	10	416.2	16.7	6405	5	AAF97850	Aaf97850 Human neu
c	11	416	16.6	126990	12	ADP13332	Adp13332 Renal cel
c	12	414.2	16.6	196686	11	ACN44170	Acn44170 Human gen
c	13	413.6	16.6	86000	12	ADP68568	Adp68568 Human PPA
c	14	413.6	16.6	215974	12	ADQ97523	Adq97523 Human can
c	15	413.4	16.5	13670	6	AAL42369	Aal42369 Human Gua
	16	413.2	16.5	6519	5	ABA16909	Aba16909 Human ner
	17	408.8	16.4	17245	4	AAK83897	Aak83897 Human imm
	18	406.2	16.3	348101	12	ADQ97146	Adq97146 Human can
c	19	405.8	16.2	227246	13	ABD33272	Abd33272 Human can
c	20	404	16.2	95240	10	ADL13556	Adl13556 Osteoarth
c	21	403.6	16.2	23639	6	ABQ78991	Abq78991 Human pho
c	22	400.4	16.0	14176	4	AAS26670	Aas26670 Human gen
c	23	400.4	16.0	14176	8	ABX74019	Abx74019 Human nov
c	24	400.4	16.0	177531	8	ACF62732	Acf62732 Cancer ba
c	25	400.4	16.0	177531	8	ADB20847	Adb20847 MRP1 base
c	26	400.4	16.0	177531	10	ADB87936	Adb87936 Human UGT
c	27	400.4	16.0	177531	10	ADB96919	Adb96919 Human MDR
c	28	400.4	16.0	177531	10	ADB92110	Adb92110 Human MDR
c	29	400.4	16.0	177531	10	ADH74617	Adh74617 Human BAC
	30	399	16.0	87687	11	ACN45166	Acn45166 Human gen
	31	399	16.0	131078	14	ADX06911	Adx06911 Cyclin-de
	32	398.6	16.0	174448	11	ACN43946	Acn43946 Human gen
c	33	398.4	15.9	167343	6	ABL64403	Abl64403 Stomach c
c	34	398.4	15.9	167343	6	ABL67239	Abl67239 Thyroid c
	35	394.8	15.8	30620	4	AAK66931	Aak66931 Human imm
	36	394.6	15.8	19300	12	ADP74371	Adp74371 Human X c
	37	392.4	15.7	9289	10	ADE84033	Ade84033 5' regula
	38	391.4	15.7	226475	9	AAD58279	Aad58279 Human tum
c	39	390	15.6	13224	4	AAS41751	Aas41751 Genomic s
c	40	390	15.6	13224	4	ABA06811	Aba06811 Human gen
c	41	390	15.6	13224	6	ABV84148	Abv84148 Human pol
	42	390	15.6	36221	4	AAS00624	Aas00624 Human dea
c	43	389.6	15.6	23456	13	ABD33110	Abd33110 Human can
	44	389.6	15.6	60057	11	ACN44314	Acn44314 Human gen
	45	388	15.5	75252	11	ACN44450	Acn44450 Human gen

ALIGNMENTS

RESULT 1
ADI35082
ID ADI35082 standard; DNA; 12174 BP.
XX
AC ADI35082;
XX
DT 22-APR-2004 (first entry)
XX
DE Human PLA2G1B nucleotide sequence.
XX
KW PLA2G1B ; fat deposition; leanness; polymorphism;
KW non-insulin dependent diabetes mellitus; NIDDM; hyperinsulinemia;
KW hypertension; glucose intolerance; dyslipidemia; hypercoagulability;
KW microalbuminuria; human; gene; ds.
XX
OS Homo sapiens.
XX
PN WO2004002295-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020830.
XX
PR 27-JUN-2002; 2002US-0392361P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML;
XX
DR WPI; 2004-082843/08.
DR P-PSDB; ADI35083.
XX
PT Diagnosing a predisposition to fat deposition or leanness, useful for
PT diagnosing a predisposition to e.g. diabetes or hypertension, comprises
PT detecting the presence of a polymorphism in the PLA2G1B nucleic acid from
PT the subject.
XX
PS Claim 1; SEQ ID NO 1; 91pp; English.
XX
CC The invention relates to diagnosing a predisposition to fat deposition or
CC leanness in a subject comprising detecting the presence or absence of a
CC polymorphic variation associated with fat deposition at a polymorphic

CC site in a PLA2G1B nucleotide sequence in a nucleic acid sample from a
CC subject, where the presence of the polymorphic variation indicates a
CC predisposition to fat deposition in the subject. The polymorphic
CC variation is a guanine at position 7328 or thymine at position 9182 of
CC the present sequence. The method is useful for diagnosing a
CC predisposition to fat deposition or leanness in a subject, and
CC consequently for diagnosing a predisposition to non-insulin dependent
CC diabetes mellitus (NIDDM) in a subject and conditions such as
CC hyperinsulinemia, hypertension, glucose intolerance, dyslipidemia,
CC hypercoagulability, or microalbuminuria, which can lead to early
CC prescription of preventive measures. The present sequence represents a
CC human PLA2G1B nucleotide sequence.
XX
SQ Sequence 12174 BP; 3217 A; 2992 C; 2738 G; 3215 T; 0 U; 12 Other;

Query Match 100.0%; Score 2499; DB 12; Length 12174;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 2500; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

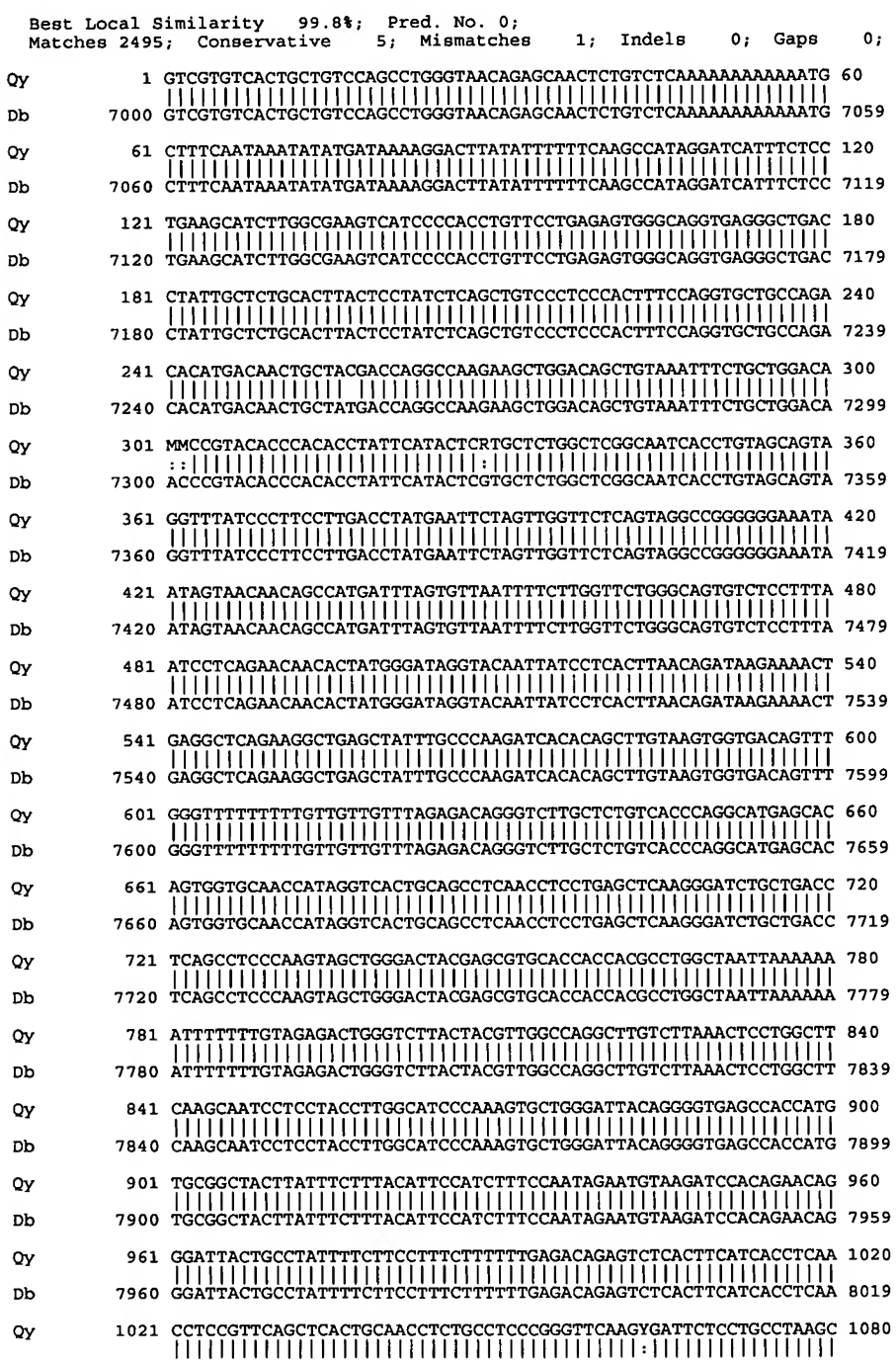
```
Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7000 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 7059

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7060 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 7119

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7120 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 7179

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7180 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 7239

Qy        241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              ||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||||
Db       7240 CACATGACAACTGCTAYGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 7299

Qy        301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7300 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 7359

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7360 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGGAAATA 7419

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7420 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 7479

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7480 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7539

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7540 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7599

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7600 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7659

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7660 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7719

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7720 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7779

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7780 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7839

Qy         841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7840 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7899

Qy         901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7900 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7959

Qy         961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7960 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 8019

Qy        1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8020 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 8079

Qy        1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8080 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 8139

Qy        1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8140 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 8199

Qy        1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8200 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 8259

Qy        1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8260 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 8319

Qy        1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8320 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 8379

Qy        1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8380 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 8439

Qy        1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8440 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8499

Qy        1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8500 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8559

Qy        1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8560 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8619

Qy        1621 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8620 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8679

Qy        1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8680 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8739

Qy        1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8740 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8799

Qy        1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8800 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8859

Qy        1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8860 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8919
```

```
Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8920 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8979

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8980 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 9039

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9040 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 9099

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9100 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 9159

Qy       2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9160 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 9219

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9220 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 9279

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9280 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 9339

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9340 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 9399

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9400 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 9459

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db       9460 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9500
```

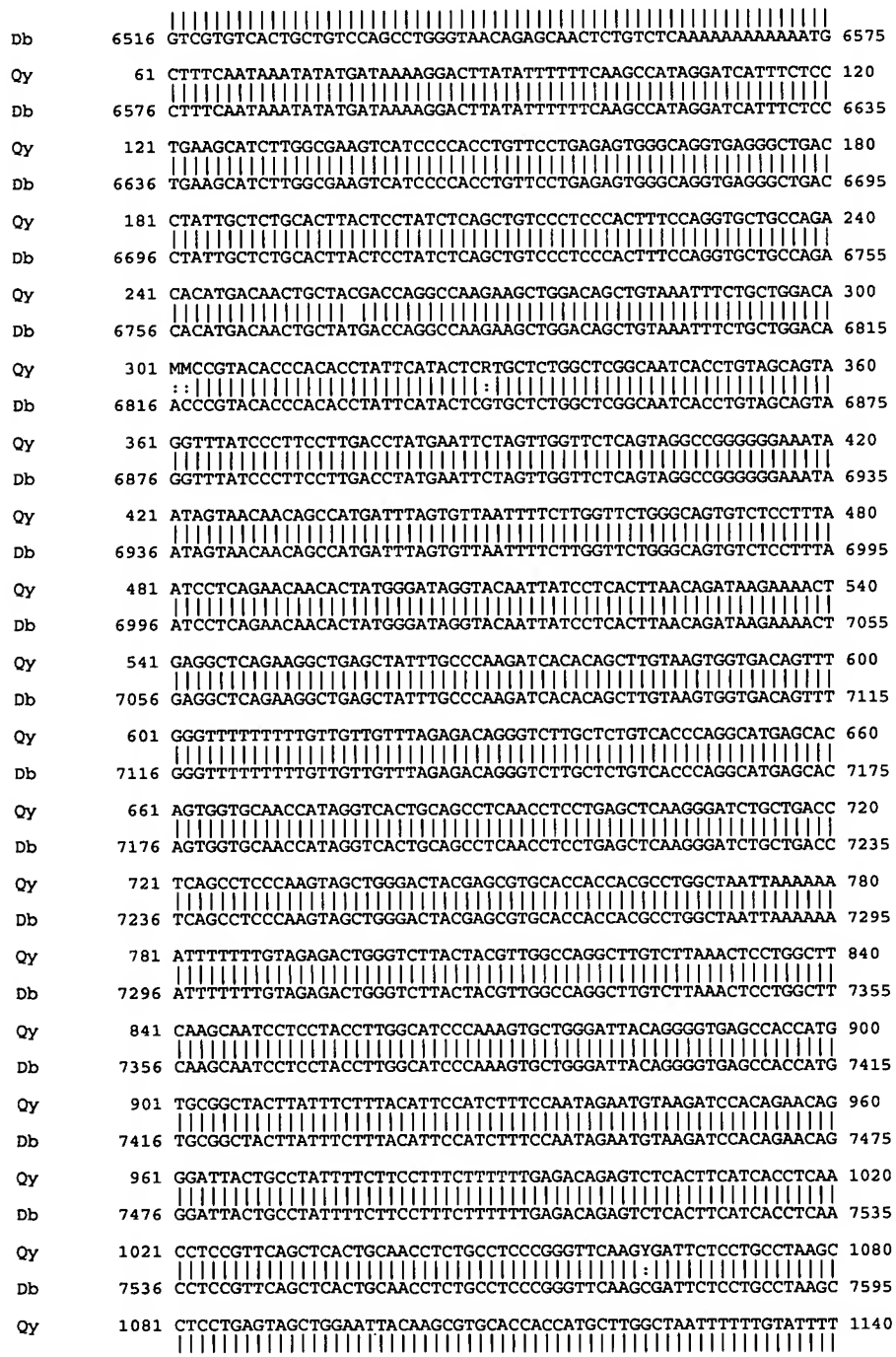
RESULT 2
ADJ09983
ID ADJ09983 standard; DNA; 12174 BP.
XX
AC ADJ09983;
XX
DT 17-JUN-2004 (first entry)
XX
DE Human phopholipase A2 (PLA2G1B) DNA SeqID 1.
XX
KW human; gene; ds; fat reduction; fat deposition; phospholipase A2;
KW PLA2G1B; chromosome 12q24; single nucleotide polymorphism; SNP;
KW appetite suppressant; lipase inhibitor; exercise regimen; obesity;
KW non-insulin dependent diabetes mellitus; NIDDM; cardiovascular disorder;
KW hypertension; antidiabetic.
XX
OS Homo sapiens.
XX
FH Key   Location/Qualifiers
FT variation   replace(436,c)
FT   /*tag= a
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(839,a)
FT   /*tag= b
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(4050,a)
FT   /*tag= c
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(4689,t)
FT   /*tag= d
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(6282,a)
FT   /*tag= e
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(6358,c)
FT   /*tag= f
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(6653,t)
FT   /*tag= g
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(7256,t)
FT   /*tag= h
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(7300,a)
FT   /*tag= i
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(7301,c)
FT   /*tag= j
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(7328,g)
FT   /*tag= k
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(8062,c)
FT   /*tag= l
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(9182,t)
FT   /*tag= m
FT   /standard_name= "Single nucleotide polymorphisms"
FT variation   replace(11649,c)
FT   /*tag= n
FT   /standard_name= "Single nucleotide polymorphisms"
XX
PN WO2004002296-A2.
XX
PD 08-JAN-2004.
XX
PF 27-JUN-2003; 2003WO-US020831.
XX
PR 27-JUN-2002; 2002US-0392362P.
XX
PA (SEQU-) SEQUENOM INC.
XX
PI Adam GIR, Langdown ML, Denissenko MF, Dennis E, Cantor C;
PI Rubin B;
XX
DR WPI; 2004-071944/07.
DR P-PSDB; ADJ09984.
XX
PT Identifying a candidate therapeutic for fat reduction, useful for
PT treating diabetes, by introducing a test molecule to a system comprising
PT PLA2G1B protein or nucleic acid, and determining the presence of
PT interaction between the compounds.
XX
PS Claim 1; SEQ ID NO 1; 116pp; English.
XX
CC This invention relates to a novel candidate therapeutic agent useful for
CC fat reduction and disorders related to fat depositions. Specifically, it
CC refers to polymorphic variations in the phospholipase A2 (PLA2G1B) DNA,
CC which is located on chromosome 12q24 and has been associated with central
CC fat deposition. The present invention describes methods to detect the
CC presence or absence of these single nucleotide polymorphisms of PLA2G1B,
CC in particular G7328A and T9182G, and subsequently provide treatment that
CC reduces fat deposition. This treatment may consist of an appetite
CC suppressant, a lipase inhibitor, a phospholipase inhibitor, an exercise
CC regimen, a dietary regimen, psychological counselling, psychotherapy or a
CC psychotherapeutic. Accordingly, PLA2G1B is a target for reducing fat
CC deposition and it can be used to treat both obesity and non-insulin
CC dependent diabetes mellitus (NIDDM), as well as cardiovascular disorders
CC such as hypertension. As such, it exhibits antidiabetic activity. This
CC polynucleotide sequence is the human PLA2G1B DNA of the invention.
XX
SQ Sequence 12174 BP; 3220 A; 2996 C; 2739 G; 3219 T; 0 U; 0 Other;

Query Match 99.9%; Score 2497.4; DB 12; Length 12174;

Best Local Similarity 99.8%; Pred. No. 0;
Matches 2495; Conservative 5; Mismatches 1; Indels 0; Gaps 0;



```
Qy       2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
             ||||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||
Db       9160 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 9219

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9220 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 9279

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9280 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 9339

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9340 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 9399

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       9400 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 9459

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db       9460 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9500


RESULT 3
ABK47376
ID   ABK47376 standard; DNA; 13612 BP.
XX
AC   ABK47376;
XX
DT   18-JUN-2002  (first entry)
XX
DE   Human Phospholipase A2, groupIB (PLA2G1B) gene.
XX
KW   Human; ds; gene; SNP; single nucleotide polymorphism; pancreatitis;
KW   pancreatic cancer; Phospholipase A2 groupIB; PLA2G1B; gene therapy;
KW   haplotype; genotype; chromosome 12q23-q24.1; transgenic; drug screening.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   variation       replace(3845,A)
FT                   /*tag=  a
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   variation       replace(3968,A)
FT                   /*tag=  b
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   CDS             4053. .9613
FT                   /*tag=  c
FT                   /product= "Pla2G1B"
FT   exon            4053. .4086
FT                   /*tag=  d
FT                   /number= 1
FT   intron          4087. .5785
FT                   /*tag=  e
FT                   /number= 1
FT   exon            5786. .5945
FT                   /*tag=  f
FT                   /number= 2
FT   intron          5946. .6744
FT                   /*tag=  g
FT                   /number= 2
FT   variation       replace(6060,A)
FT                   /*tag=  h
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   exon            6745. .6873
FT                   /*tag=  i
FT                   /number= 3
FT   variation       replace(6844,A)
FT                   /*tag=  j
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
FT   intron          6874. .9489
FT                   /*tag=  k
FT                   /number= 3
FT   exon            9490. .9613
FT                   /*tag=  l
FT                   /number= 4
FT   variation       replace(9531,A)
FT                   /*tag=  m
FT                   /label= SNP
FT                   /note= "Single nucleotide polymorphism"
XX
PN   WO200212562-A2.
XX
PD   14-FEB-2002.
XX
PF   06-AUG-2001; 2001WO-US024663.
XX
PR   04-AUG-2000; 2000US-0223179P.
XX
PA   (GENA-) GENAISSANCE PHARM INC.
XX
PI   Kazemi A,  Kliem SE,  Koshy B;
XX
DR   WPI; 2002-303982/34.
DR   P-PSDB; AAU78667.
XX
PT   Novel isolated human Phospholipase A2, Group IB pancreas polynucleotide,
PT   for therapeutic purposes, for studying expression and function of the
PT   polynucleotide and for expressing the phospholipase protein.
XX
PS   Claim 1; Fig 1; 51pp; English.
XX
CC   The invention relates to an isolated human Phospholipase A2, Group IB
CC   (pancreas) (PLA2G1B) polynucleotide comprising a sequence which is a
CC   polymorphic variant for a reference sequence for the PLA2G1B gene or its
CC   fragment, or a polymorphic variant of a reference sequence for a PLA2G1B
CC   cDNA or its fragment. Also included are haplotyping/genotyping the
CC   PLA2G1B gene of an individual, predicting the haplotype pair for the
CC   PLA2G1B gene of an individual, identifying an association between a trait
CC   and at least one haplotype or haplotype pair of the PLA2G1B gene, an
CC   isolated genotyping oligonucleotide for detecting a polymorphism in the
CC   PLA2G1B gene, a recombinant non-human organism transformed or transfected
CC   with the PLA2G1B sequence, where the organism expresses a PLA2G1B protein
CC   encoded by the first nucleotide sequence or by the polymorphic variant
CC   sequence, an isolated polypeptide comprising a sequence which is a
CC   polymorphic variant of a reference sequence for the PLA2G1B protein or
CC   its fragment, an anti-PLA2G1B monoclonal antibody, screening for drugs
CC   targeting PLA2G1B, a computer system for storing and analysing
CC   polymorphism data for the PLA2G1B gene and a genome anthology for PLA2G1B
CC   gene. The PLA2G1B variant is useful in studying the expression and
CC   function of PLA2G1B, and in expressing PLA2G1B protein for use in
CC   screening for candidate drugs to treat diseases related to PLA2G1B
CC   activity (e.g. pancreatitis and pancreatic cancer) and for therapeutic
CC   purposes. The transgenic organism is useful for studying expression of
CC   the PLA2G1B isogenes in vivo, for in vivo screening and testing of drugs
CC   targeted against PLA2G1B protein, and for testing the efficacy of
CC   therapeutic agents and compounds in a biological system. The antibody is
CC   useful for studying the effect of the variation on the biological
CC   activity of PLA2G1B as well as on the binding affinity of candidate drugs
CC   targeting PLA2G1B. The present sequence is the PLA2G1B gene which is
CC   located on chromosome 12q23-q24.1
XX
SQ   Sequence 13612 BP; 3637 A; 3290 C; 3070 G; 3615 T; 0 U; 0 Other;

  Query Match             99.9%;  Score 2497.4;  DB 6;  Length 13612;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2495;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
```


Db 6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575
Qy 61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
Db 6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635
Qy 121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
Db 6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695
Qy 181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
Db 6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755
Qy 241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
Db 6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815
Qy 301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
Db 6816 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875
Qy 361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
Db 6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935
Qy 421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
Db 6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995
Qy 481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
Db 6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055
Qy 541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
Db 7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115
Qy 601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
Db 7116 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175
Qy 661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
Db 7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7235
Qy 721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
Db 7236 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7295
Qy 781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
Db 7296 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7355
Qy 841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
Db 7356 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7415
Qy 901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
Db 7416 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7475
Qy 961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
Db 7476 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 7535
Qy 1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db 7536 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 7595
Qy 1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db 7596 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 7655
Qy 1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db 7656 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 7715
Qy 1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db 7716 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 7775
Qy 1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db 7776 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 7835
Qy 1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db 7836 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 7895
Qy 1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db 7896 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 7955
Qy 1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
Db 7956 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8015
Qy 1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
Db 8016 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8075
Qy 1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
Db 8076 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8135
Qy 1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
Db 8136 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8195
Qy 1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
Db 8196 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8255
Qy 1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
Db 8256 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8315
Qy 1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
Db 8316 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8375
Qy 1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
Db 8376 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8435
Qy 1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
Db 8436 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8495
Qy 1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
Db 8496 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 8555
Qy 2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
Db 8556 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 8615
Qy 2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
Db 8616 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 8675
Qy 2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
Db 8676 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 8735

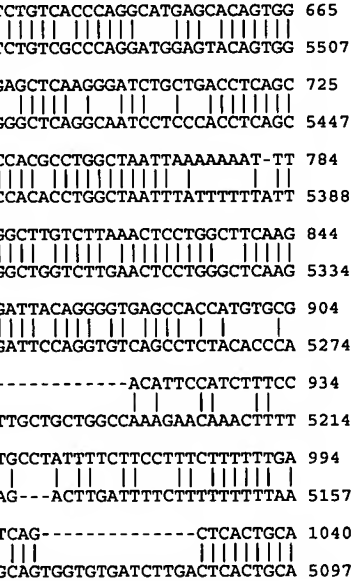
```
Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8736 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 8795

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8796 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 8855

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8856 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 8915

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       8916 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 8975

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db       8976 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9016


RESULT 4
ABD33466/c
ID   ABD33466 standard; DNA; 42360 BP.
XX
AC   ABD33466;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-089.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 618; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX

SQ   Sequence 42360 BP; 11698 A; 9658 C; 10445 G; 10559 T; 0 U; 0 Other;

  Query Match             18.3%;  Score 456.6;  DB 13;  Length 42360;
  Best Local Similarity   99.8%;  Pred. No. 4.8e-84;
  Matches  456;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy       2045 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 2104
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42360 ACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAAAGTT 42301

Qy       2105 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 2164
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42300 TAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAA 42241

Qy       2165 AGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 2224
             ||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||||||
Db      42240 AGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCTATAA 42181

Qy       2225 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 2284
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42180 TCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACACTAGC 42121

Qy       2285 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 2344
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42120 TTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCTATGT 42061

Qy       2345 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 2404
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42060 AGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGAAACT 42001

Qy       2405 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 2464
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      42000 AGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGA 41941

Qy       2465 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||
Db      41940 TCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 41904


RESULT 5
ABD33504
ID   ABD33504 standard; DNA; 93544 BP.
XX
AC   ABD33504;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-098.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 676; 182pp; English.
XX
```

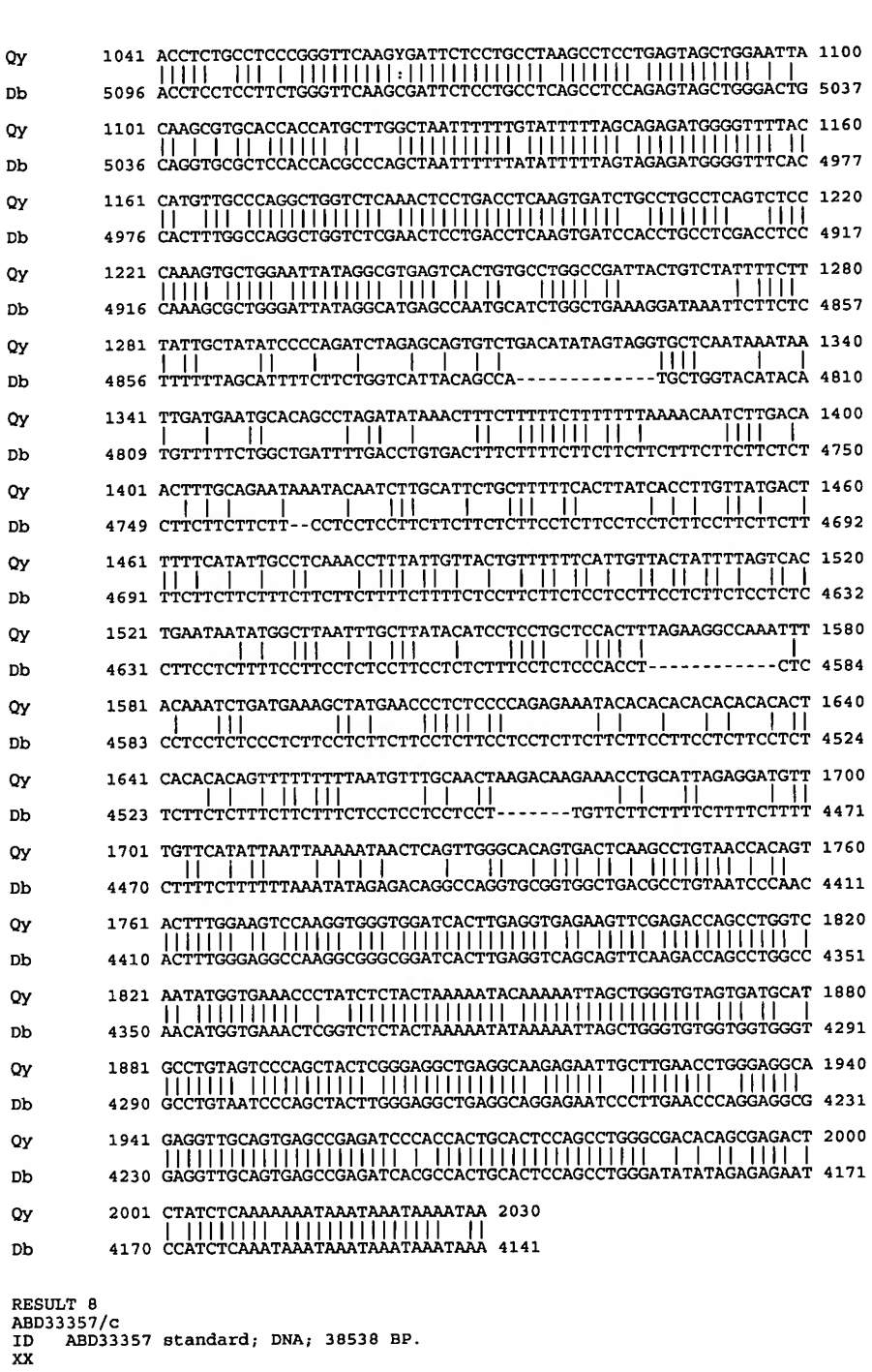
```
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 93544 BP; 26138 A; 20027 C; 21635 G; 25744 T; 0 U; 0 Other;

  Query Match             17.4%;  Score 434.2;  DB 13;  Length 93544;
  Best Local Similarity   59.7%;  Pred. No. 2.4e-79;
  Matches  861;  Conservative    1;  Mismatches  559;  Indels   22;  Gaps    7;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
Db      18099 TTTTTTTTTTTTTTTTTGAGAGATGGAGTCTTAGCCTGTTGCCCAGGCTGGAGTGCAGT 18158

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
Db      18159 GGCACCATCTTGGGTCACTGCAACCTCCACCATGTGGGTTCAAGAGATTCTACTGCCTCA 18218

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db      18219 ACCTCCCAAGTAGCTGGAATTACAGGCTTGTGCCACCACACCTTGCTAATT--TTTGTGT 18276

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db      18277 TTTTATTAGAGACAGGGTTTCGCCATGTTGGCCAGGCTGATCTCAAACTCCTGGCCTCAA 18336

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db      18337 CTGATCCGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAAGTGTGAGCCACCATGCCT 18396

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db      18397 GGCCACCACAGGCTTTATTTACAGGTTGCCTCATTACCTGAATTCTCTCTCTCCAAATTC 18456

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA---A 1020
Db      18457 TTGTTACCCAAACACAGGAACATGAGATGGAGTCTCGCTCCGTCGCCCAGTCTGGAGTGC 18516

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db      18517 AGTGATCTCAGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGTGATTCTTCTGCCTCAGC 18576

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db      18577 CTCCCAAGTAGCTGGGACTACAGGCATGCGCCACCCCACCTGGCTAA-TTTTTGTATTTT 18635

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db      18636 TAGTAGAGACGGGGTTTCACCATGCTGGCCAGGCTGGTCTCGAACTCCTGACCTC--GTG 18693

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db      18694 ATCCACCCGCCTCAGCTTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCATACCTGGC 18753

Qy       1261 CGATTACTGTCTATTTT------CTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314
Db      18754 CAGGAACTAGACAGTTTTAGAGACATCTTTCTGAAATAGGGCTGGTGCCCCTCATTATTC 18813

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374
Db      18814 ATTAATGCTATTTTCTCCAAAACCCGGGAACCAAATAGATGTGATAGCATGGCATCCCTC 18873

Qy       1375 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434
Db      18874 ATGATCTGAAGTTACTGTCCGTAACTGCGGTCCCATTCGATTTGGATATTGGGGGATGCC 18933

Qy       1435 TTTTCACTTATCA-CCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTG 1493
Db      18934 TGCACGTGTCTGTGTCATGTGCTGCATGTCTGCGGTGCCATCAGTCTGTGAGCTCCCTCT 18993

Qy       1494 TTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCT 1553
Db      18994 GGGCCTGCGTCCGTTTTTCTCATCCTTTGTATTATCTCAGCACTTTGTACAATACCTGGC 19053

Qy       1554 CCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCC 1613
Db      19054 ACTCAATGAATTAAAAATAATCTGAAGCGGGGGTTGGGGGGAGCTTGTCAATATAGCCCC 19113

Qy       1614 AGAGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAA 1673
Db      19114 ACCTTGTTTTGCCCTTGGACTTGATAAAGCCGTTTGTTGCTATGCATATGTCCATCTCTT 19173

Qy       1674 GACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCA------- 1726
Db      19174 TTCTAGACAATGAGTTACTTGGATGAATAGGCAGATCTCTTGGTTAAACCATAAAGACAG 19233

Qy       1727 GTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGAT 1786
Db      19234 GCTGGGCGCGGTGGCTCACATCTGTAATCCCAGCACTTTGAGAGGCCGAGGCGGGCAGAT 19293

Qy       1787 CACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTA 1846
Db      19294 CACTTGAGGTTAGAAGTTCGATACCAGCCTGGCCAACATGGTGAAACCCAATCTCTACTA 19353

Qy       1847 AAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGG 1906
Db      19354 AAAATACAAAAATTAGCCGAGTGCGGTGGCGCATGCCTGTAATCCTAGCTACTTGGGAGG 19413

Qy       1907 CTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCAC 1966
Db      19414 CTGAGGCAGGAGAATCGCTTGAACCCGGGAGGTGAAGCTTTCAGTGAGCCGAGATTGTGC 19473

Qy       1967 CACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAA 2026
Db      19474 CATTACACTCCAGCCTGGGCAACAGAGCGAGACTGTGTCTCAAAATGAAAACCCATAAAG 19533

Qy       2027 ATA 2029
Db      19534 ACA 19536


RESULT 6
ACN43914/c
ID   ACN43914 standard; DNA; 160482 BP.
XX
AC   ACN43914;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG26773.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
```

```
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 100; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 160482 BP; 44060 A; 32143 C; 33530 G; 49875 T; 0 U; 874 Other;

  Query Match             17.4%;  Score 434.2;  DB 11;  Length 160482;
  Best Local Similarity   60.2%;  Pred. No. 2.7e-79;
  Matches  888;  Conservative    1;  Mismatches  525;  Indels   61;  Gaps    8;

Qy        606 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 665
Db       5181 TGTTTTTTTTTTTCTCTCGATACGGAGTCTTGCTCTGTTGCCCAGGCTGGAGTGCAGTGG 5122

Qy        666 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 725
Db       5121 CATGATCTCGGCTCACTGAAACCTCTGCCTCCTGGGTTCACGTGATTCTCCTGCCTCAGA 5062

Qy        726 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGG--CTAATTAAAAAAATT 783
Db       5061 CTCCCAAGTAGCTGGGATTACAGGCGTGAGCCACCATGCCTGGCCCTAATGTTTTGTATT 5002

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db       5001 TTTTAGTAGAGAC-AGGTTTCGCTACATTGGTCAGGCTAGTCTCAAACTCCTGACCTCAA 4943

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db       4942 GTGATCCGCCTGCCTTGATGTACCAAAGTGCTGGGATTACAGATGTGAGCCACCCCGCCC 4883

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db       4882 GGCCAGTCATTCACTTTCTTTCTTTTTTTTTTTTTTTTGAGACAGAGTCTCATTCTGTC 4823

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
Db       4822 ATCCAGGCTGGAGTGT----------------------------------AGTGGCAC 4799

Qy       1024 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 1083
Db       4798 GATCTCGGCTCACTTCAACTTTCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTC 4739

Qy       1084 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAG 1143
Db       4738 TTGAGTAGCTGGGATTACAGGCATGCACCACCGCACCCGGCTAATTTTTTATATTTTTGG 4679

Qy       1144 CAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATC 1203
Db       4678 GAGAGATGGGGTTTCACCATGTTAGACAGGCTGGTCTCAAACTCCTGACCTCAAGCAATC 4619

Qy       1204 TGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGA 1263
Db       4618 TGCTCGCCTTGGCCTCCTAAAGTGCTGGGATTACAGGCGTGAGCCACCACACCCGGCCAA 4559

Qy       1264 TTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGT 1323
Db       4558 --ATCATTCACTTTCTACAACAAGTTCTGTCTGTATGCCAATGAATCTCAGGACTGTACT 4501

Qy       1324 AGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTT 1383
Db       4500 TCTAGCTCTAACCTCTAAACATATATTTAAGTGTCCTTTAAAGCCAGGACTTAGATTCCT 4441

Qy       1384 TTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTT 1443
Db       4440 CAGAATCTCAAATTCAACAAATCCAAAATTCAACATATATTTTCCCTAAATTCTGCCCCT 4381

Qy       1444 ATCACCTTGTTA------TGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTT 1497
Db       4380 TTCTCTGTATTTCCTAGTTTGCTTAATAGTATCACCTTAATTCAAGGAAGAAATTTCAAA 4321

Qy       1498 TTCATTGTTACTATTTTAGTCACTGAAT--AATATGGCTTAATTTGCTTATACATCCTCC 1555
Db       4320 GTCATCTTCAATTTTTCTCTCTCTCAACGCAATAAAGCACTAATTCCTAGAAGCTCAACC 4261

Qy       1556 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1615
Db       4260 TTAAACAGTCCTCTGGAATGAGATGAAGGTTAACAGTTACTGAACACAGACCAAAACTAG 4201

Qy       1616 AGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGA 1675
Db       4200 GCACAGCAGCAAGGGTTGTTCAATTGAATTCAATTCTCTCAGAAGCCCTAAGGAATAGCT 4141

Qy       1676 CAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAAT------AACTCAGTT 1729
Db       4140 CCAATTTTATAGAGGAAGAAACTGAGGTTCCAAGAAGTAAGAAATTCACACAAGGAGGCT 4081

Qy       1730 GGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCAC 1789
Db       4080 GGGTGTGGTGGCTCACGCCTGTAATCCTAGCACTTTGGGAGGCCAAGGCGGGCAGATCAC 4021

Qy       1790 TTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAA 1849
Db       4020 CTGAGGTCAAGAGCTCGAGACCAGCCTGACAAACATGGTGAAACCCCATCTCTACTAAAA 3961

Qy       1850 ATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTG 1909
Db       3960 ATACAAAAATTACCCAGGTGTGGTGGTGCACGCCTGTAATCCCAGCTATTTGGGAGGCTG 3901

Qy       1910 AGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCC------GAGATCC 1963
Db       3900 ACGCAGGAGAATTGTTTGAACCCAGGAGGCGGAGGTTGCAGTGAGCCAAGATGGAGATCG 3841

Qy       1964 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 2023
Db       3840 TGCCACTGCACTCCAGCCTGGGCAACAGAGCGAGACTCTGTCTCCAAAAAAAAAAAAAAA 3781

Qy       2024 AAAATAAAGGATCGGAGAGAAACAAAACTAATAAG 2058
Db       3780 AGAAAGAAAGAAAGAAAAAAATAAATTCACACAAG 3746


RESULT 7
ACN44410/c
ID   ACN44410 standard; DNA; 91760 BP.
XX
AC   ACN44410;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG41365.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
```

XX
PN WO2003073826-A2.
XX
PD 12-SEP-2003.
XX
PF 28-FEB-2003; 2003WO-US006235.
XX
PR 01-MAR-2002; 2002US-00087192.
XX
PA (SAGR-) SAGRES DISCOVERY.
XX
PI Morris DW;
XX
DR WPI; 2003-328604/31.
XX
PT Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT comprises a nucleotide sequence.
XX
PS Claim 1; SEQ ID NO 844; 0pp; English.
XX
CC The present invention relates to novel DNA and protein sequences which
CC are associated with carcinomas. The sequences are useful for: (i) for
CC screening drug candidates; (ii) for screening of bioactive agent capable
CC of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC a bioactive agent capable of modulating the activity of CAP; (iv) for
CC evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC determining Carcinoma Associated (CA) gene copy number. In addition, the
CC CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC carcinoma including lymphoma. The present sequence is one such CA coding
CC sequence. Note: This patent is an equivalent to basic patent
CC US2002182586A1, for which no sequence data was published
XX
SQ Sequence 91760 BP; 21177 A; 23589 C; 24129 G; 22845 T; 0 U; 20 Other;

Query Match 17.3%; Score 433.2; DB 11; Length 91760;
Best Local Similarity 61.6%; Pred. No. 3.8e-79;
Matches 906; Conservative 1; Mismatches 474; Indels 89; Gaps 10;





RESULT 8
ABD33357/c
ID ABD33357 standard; DNA; 38538 BP.
XX

```
AC   ABD33357;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human cancer-associated (CA) gene HD07-065.
XX
KW   Human; cancer-associated protein; CAP; cancer-associated gene; CA; gene;
KW   ds; cancer; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO2004058146-A2.
XX
PD   15-JUL-2004.
XX
PF   15-DEC-2003; 2003WO-US040081.
XX
PR   17-DEC-2002; 2002US-00322281.
XX
PA   (SAGR-) SAGRES DISCOVERY INC.
XX
PI   Morris DW,  Malandro MS;
XX
DR   WPI; 2004-499109/47.
XX
PT   Novel human cancer associated protein encoded within open reading frame
PT   of cancer associated gene, useful as targets for diagnosing cancer.
XX
PS   Claim 16; SEQ ID NO 448; 182pp; English.
XX
CC   The invention relates to cancer-associated proteins (CAP) and the cancer-
CC   associated (CA) nucleic acids encoding them. The invention also relates
CC   to a method for treating cancers involving administering to a patient an
CC   inhibitor of CAP, and a method of screening for anticancer activity in a
CC   potential drug involving providing a cell that expresses a CA gene,
CC   contacting a tissue sample derived from a cancer cell with an anticancer
CC   drug candidate and monitoring the effect of the anticancer drug candidate
CC   on expression of the CA gene. The CAP proteins are useful for detecting
CC   cancer associated with expression of a CAP protein in a test cell sample
CC   and for screening for a bioactive agent capable of modulating the
CC   activity of a CAP protein. The CA nucleic acids are useful for diagnosing
CC   cancer, involving determining the expression of a CA nucleic acid in a
CC   tissue. This sequence represents a human CA gene of the invention. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 38538 BP; 9418 A; 9893 C; 9523 G; 9684 T; 0 U; 20 Other;

  Query Match             17.0%;  Score 425.4;  DB 13;  Length 38538;
  Best Local Similarity   60.0%;  Pred. No. 1.3e-77;
  Matches  879;  Conservative    1;  Mismatches  552;  Indels   33;  Gaps    9;

Qy        567 CCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAG 626
Db      35304 CAAAAAATAAAAATTAAATATTTTTTGTTTCTTAGATTATAGTTAAGGGGCTTTTTTGAG 35245

Qy        627 ACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCA 685
Db      35244 ATGAAGTTTTGCTCTTGTTGCCCAGGCTGGAGTGCAATGGTGTGATCTTGGCTCACAGCA 35185

Qy        686 GCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTA 745
Db      35184 ACCTCCACCTCCAGGGTTCAAGCGATTCTCCTCTCTCAGCCTCCTGAGTAGCTGGGATTA 35125

Qy        746 CGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTA 805
Db      35124 CAGGCATGCGCCACCACACCCAGCTAATTTTGTA---TTTTTAGTAGAGATGGGGTTTCT 35068

Qy        806 CTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATC 865
Db      35067 CCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGGTGATCCACCCTCCTTGGCTTC 35008

Qy        866 CCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC 925
Db      35007 CCGAAGTGTTGGGATTACAGGCATGAGCCACTGCGCCTGGC------CTCAGTTAAGGGA 34954

Qy        926 CATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTT 985
Db      34953 CTTTTTACATATTTATTTATTTATTTATTTATTATTATTTTGAGACGAAGTCTCGCTGTT 34894

Qy        986 CTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTC 1045
Db      34893 GTTCTCCAGGCTGGAGTGCAATGGCACGACC-----------TTGGCTCACTGCAACCTC 34845

Qy       1046 TGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGC 1105
Db      34844 CGCCTCCCGGGTTCAAGCGATTCTCCTGTCTTGGCCTCCCAAGTAGCTGGGATTACAGGC 34785

Qy       1106 GTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGT 1165
Db      34784 GCCTGCCATCATGCCTGGCTAA-TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATTA 34726

Qy       1166 TGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAG 1225
Db      34725 TGGCCAGGCTGGTCTTGAAGTCCTGACTTCAGGTGATCTGCCTGCCTCGGCCTCCCAAAG 34666

Qy       1226 TGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTG 1285
Db      34665 TGCTGGGATGACAGGTGTGAGCCACCACACTCGGCCTTTATTTATTATTTTCGAGACAGG 34606

Qy       1286 CTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGAT 1345
Db      34605 GTATCTC------TCTGTTGCCCAGGCTGGAGTGCAGTGGTGCAGTCACGGCTGGTTGCA 34552

Qy       1346 GAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTT 1405
Db      34551 GCCTCAAACTCCTGGGCTCCAGCA--ATTCTCCTGCCTCAGCCTCCTGGGTAGCTGGAAC 34494

Qy       1406 GCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTC 1465
Db      34493 TACAGGTGCAAATTATCATGCCTGGCTAATTTTTAAATTTTTTGTCAAGATGGGGTCTCT 34434

Qy       1466 ATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAAT 1525
Db      34433 CTATATTGCCCAGGCTGGTCTCAAACTCCTGGCCTCAAGTGATCTCCCTTCTGGGCTTCA 34374

Qy       1526 AATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAA 1585
Db      34373 AAAAGTGCTGGGATTATAAGTATGAGCCACCACCTCCGGCCTTCCCTCCTTCCTTCCTTC 34314

Qy       1586 TCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACAC 1645
Db      34313 CTTTCTGCCTTCCTTCCTTCCTTTCTCAAAAAAAAAAAAAAAAGAAGTCAGTGTTGTCTGA 34254

Qy       1646 ACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTC 1705
Db      34253 CTGGTATTGTCTGATTACCTGCATATATTGATATATTATTCCAATAAAGATGCTTGCTTT 34194

Qy       1706 ATA--TTAATTAAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACT 1763
Db      34193 TTCCTTTGGTTAAAAAGTATCCGGCCAGGCGCGATGGCTCACGCTTGTAACCCCAGTACT 34134

Qy       1764 TTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAAT 1823
Db      34133 TTGGGAGGCTGTGGTGGGCGGATCACCTGAGATCAGGAGTACGAGACCAGCCTGGCCAAC 34074

Qy       1824 ATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCC 1883
Db      34073 ATGGTAAAACCCCATCTCTACTAGAAATACAAAAATTAGCTGAGTGTAGTGGCACGCACC 34014

Qy       1884 TGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAG 1943
Db      34013 TGTAATCCCAGCTACTTGGGAGGCCAAGGGAGAAGAATTGTTTGAACCCGGGAGGCAGAG 33954

Qy       1944 GTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTA 2003
```

```
                ||||||||||||| ||||| ||||| |||||||||||| |||||||| || ||||||||
Db      33953 GTTGCAGTGAGCCAAGATCACACCATTGCACTCCAGCCAGGGCGACA-AGAGAGACTCTG 33895

Qy       2004 TCTCAAAAAAAATAAATAAATAAAAT 2028
              |||||||||||| ||| ||| |  ||
Db      33894 TCTCAAAAAAAAAAAAAAAAAGCAT 33870


RESULT 9
ABK83569
ID   ABK83569 standard; cDNA; 122888 BP.
XX
AC   ABK83569;
XX
DT   14-AUG-2002  (first entry)
XX
DE   Human cDNA differentially expressed in granulocytic cells #140.
XX
KW   Human; ss; granulocytic cell; DNA chip; bacterial infection;
KW   viral infection; parasitic infection; protozoal infection;
KW   fungal infection; sterile inflammatory disease; psoriasis;
KW   rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;
KW   cardiac reperfusion injury; renal reperfusion injury; ARDS;
KW   adult respiratory distress syndrome; inflammatory bowel disease;
KW   Crohn's disease; ulcerative colitis; periodontal disease;
KW   granulocyte activation; chronic inflammation; allergy.
XX
OS   Homo sapiens.
XX
PN   WO200228999-A2.
XX
PD   11-APR-2002.
XX
PF   03-OCT-2001; 2001WO-US030821.
XX
PR   03-OCT-2000; 2000US-0237189P.
XX
PA   (GENE-) GENE LOGIC INC.
XX
PI   Beazer-Barclay Y,  Weissman SM,  Yamaga S,  Vockley J;
XX
DR   WPI; 2002-435328/46.
XX
PT   Detecting granulocyte activation by detecting differential expression of
PT   genes associated with granulocyte activation, which serves as diagnostic
PT   markers that is useful for monitoring disease states and drug toxicity.
XX
PS   Claim 1; SEQ ID NO 140; 114pp; English.
XX
CC   The invention relates to detecting (M1) granulocyte (GC) activation
CC   (GCA), by detecting the level of expression of gene(s) (Gs) identified by
CC   DNA chip analysis as given in the specification, and comparing the
CC   expression level to an expression level in an unactivated GC, where
CC   differential expression of Gs is indicative of GCA. Also included are
CC   modulating (M2) GA by contacting GC with an agent that alters the
CC   expression of at least one gene in Gs; (2) screening (M3) for an agent
CC   capable of modulating GCA or an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease using the gene expression
CC   profile; (3) detecting (M4) an inflammation (especially chronic) in a
CC   tissue, an allergic response in a subject, exposure of a subject to a
CC   pathogen or sterile inflammatory disease, by detecting the level of
CC   expression in a sample of the tissue of gene(s) from Gs, where the level
CC   of expression of the gene is indicative of inflammation; (4) treating
CC   (M5) an inflammation (especially chronic) or in a tissue, an allergic
CC   response in a subject, exposure of a subject to a pathogen or sterile
CC   inflammatory disease, by contacting a tissue having inflammation with an
CC   agent that modulates the expression of gene(s) from Gs in the tissue. M1
CC   is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful
CC   for screening an agent capable of modulating GCA preferably in an
CC   inflammation in a tissue; M4 is useful for detecting an inflammation
CC   (especially chronic) in a tissue, an allergic response in a subject,
CC   exposure of a subject to a pathogen or sterile inflammatory disease (e.g.
CC   psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis,
CC   cardiac reperfusion injury, renal reperfusion injury, ARDS, adult
CC   respiratory distress syndrome, inflammatory bowel disease, Crohn's
CC   disease, ulcerative colitis, periodontal disease; also bacterial
CC   infection, viral infection, parasitic infection, protozoal infection,
CC   fungal infection and M5 is useful for treating one of the above
CC   conditions. The present sequence represents a gene differentially
CC   expressed in granulocytes. Note: The sequence data for this patent did
CC   not form part of the printed specification, but was obtained in
CC   electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 122888 BP; 28761 A; 33410 C; 31919 G; 28798 T; 0 U; 0 Other;

  Query Match             16.7%;  Score 417.2;  DB 6;  Length 122888;
  Best Local Similarity   61.1%;  Pred. No. 8.1e-76;
  Matches  894;  Conservative    0;  Mismatches  503;  Indels   65;  Gaps   11;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
              ||| |||||  | || || | | ||||| ||||||||||||||||  ||| ||||
Db      70248 TTTCTTTTTTATTTTTTTAACACACAGGTCCTTGCTTTGTCACCCAGGCTGGAGTGCAGT 70307

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
              ||||||| || || | || ||| |||||| | |||| |||||||  |||     ||||||
Db      70308 GGTGCAAACACAGCTGACAGCAACCTCAAAC-CCTGGGCTCAAGCAATCCTTCCACCTCA 70366

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAA---------- 772
              || ||||||| ||||||||||||  || ||   ||||| |||  |||||
Db      70367 GCTTCCCAAGAAGCTGGGACTACAGGCATGTGACACCATGCCCAGCTAATTTTTTTTTTT 70426

Qy        773 -----TTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTT 827
                  ||  ||||||||||||      |    | |||| ||||  |||  ||||| ||||
Db      70427 TTTACTTTTAAAAATTTTTTGTAGAAATGGGAGTCTCACTATATTGCTTAGGCTGGTCTC 70486

Qy        828 AAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGG 887
              ||| || |||  | |||||||||||||| ||| ||| |||||||||  |||||||| |||
Db      70487 AAATTCTTGGGCTTAAGCAATCCTCCTGCCTCGGCCTCCCAAAGTTGTGGGATTATAGGT 70546

Qy        888 GTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 947
              |||    ||||     | |  | |   | | ||   |       |   ||||| || | |
Db      70547 GTG----ACCACTGCAGCCAGCCTTAATTTCTAATATGGTAAGTATTAATAGCATTTTGG 70602

Qy        948 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007
                |  || |          | | | || | || ||| ||||||||||  | ||||  |
Db      70603 GGATGCTCAATAAT-----TTTCTTCTTCTTTTTTTTTTTTTTTGAGATGGGGTCTAGC 70657

Qy       1008 TTCATCACCTCAACCTCCG--------------TTCAGCTCACTGCAACCTCTGCCTCCC 1053
                ||||||  ||    |                |||||  ||||||||||||||||||||
Db      70658 ACTGTCACCTAGACTGGAGTGCAATGGCGCGATTTCAGTTCACTGCAACCTCTGCCTCCT 70717

Qy       1054 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 1113
              |||||||| ||||||| ||||| |||||||||||||||| | |||| ||   |  |||
Db      70718 GGGTTCAAGAGATTCTCGTGCCTCAGCCTCCTGAGTAGCTGGGACTACAGGCACCCGCCA 70777

Qy       1114 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 1173
              |||     |||||||||||||||||||||| |||||| ||||||  | ||||||  |||||
Db      70778 CCACATCCGGCTAATTTTTTGTATTTTTAGTAGAGATAGGGTTTCGCTATGTTGGCCAGG 70837

Qy       1174 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1233
              |||||||||||||||||||||  |||||| | ||||||||| ||||||||||||||| |
Db      70838 CTGGTCTCAAACTCCTGACCTC--GTGATCCGACTGCCTCAGCCTCCCAAAGTGCTGGGA 70895

Qy       1234 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1293
              ||| ||| ||||| |||     |  ||          | ||  | |||   | ||||
Db      70896 TTACAGGTGTGAGCCACCACACTCGG-----------CAATAATTTTTTAAGGTATAAAA 70944

Qy       1294 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1353
                 ||| ||          | |       |    | |||| |||| ||||        |
Db      70945 ATGTCCAGAGACCAAAAAGGTCAAGAACCACTGCTTTAAATACATAAATGAT-----ATC 70999

Qy       1354 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1413
              |  |||  ||||   ||||     |   |||| |||| | ||   |||         ||  |
Db      71000 AAACTATTTATATTATTTCAAAATTACTTTTAAAAAAATTCAACACAGAGCATAAG-CTC 71058
```

```
Qy       1414 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1473
Db      71059 CATGAAGCAGAGGTTTTTCTTTTTGTTTTATTCATTATGGAAACTAAGTCCCTAGAACAA 71118

Qy       1474 TCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGC 1533
Db      71119 TGCCTGGAACATAGTGGGTGTTCAATAAATACTGATTGAATGAATAAAGAATTATATTTT 71178

Qy       1534 TTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGA 1593
Db      71179 GGGTTTTACCCATTCTGATACGTGGACCTCAAACTCATTCTGCTGTTCAGTATTCTATTA 71238

Qy       1594 AAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTT 1653
Db      71239 TTCCTGAAGATGAAAAATATCACAATAACAATTCTATTAATGAACATTTGTTCATAATTTT 71298

Qy       1654 TTTTTAA----TGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1709
Db      71299 TCATGGACATTTTTATACACAGGTAAGAGTTTCTCCAGGGTATATACCTAAGAATGAAAC 71358

Qy       1710 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1769
Db      71359 TGTAGTAAATAAATCTGGCCAGGCACGGTGGCTCATGCCTGTAATCCCAGCACTTTGGGA 71418

Qy       1770 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1829
Db      71419 GGCCGAGGCAAGTGGATCAC--AAGGTCAGGGGATCAAGACCATCCTGGCCAACATGGTG 71476

Qy       1830 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1889
Db      71477 AAACCCTGTCTCTACTAAAAATACAAAAATTAGCCAGGTGTGGGGGCGTGTGCCTGTAGT 71536

Qy       1890 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1949
Db      71537 CCCAGCTACTCAGGAGGCTGAGGCAGGAGAATCGTTTGAACCCAGGAGGTGGAGGTTGCA 71596

Qy       1950 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 2009
Db      71597 GTGAGCCAAGATCGCACCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCATCTCAC 71656

Qy       2010 AAAAATAAATAAATAAAATAAA 2031
Db      71657 AAACCAAAAAGAAAAAAAAAAA 71678
```

```
RESULT 10
AAF97850/c
ID   AAF97850 standard; DNA; 6405 BP.
XX
AC   AAF97850;
XX
DT   31-MAY-2001  (first entry)
XX
DE   Human neuroblastoma cell line NB-1 1p36 nucleotide sequence SEQ ID NO:64.
XX
KW   Human; chromosome 1; 1p36; neuroblastoma cell line; NB-1; anticancer;
KW   tumour suppressor; human 1p36 homozygosity deletion domain; tumour;
KW   diagnosis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200116311-A1.
XX
PD   08-MAR-2001.
XX
PF   31-AUG-2000; 2000WO-JP005930.
XX
PR   31-AUG-1999;   99JP-00245962.
PR   09-MAY-2000; 2000JP-00136266.
XX
PA   (HISM ) HISAMITSU PHARM CO LTD.
PA   (CHIB-) CHIBA PREFECTURE.
XX
PI   Nakagawara A;
XX
DR   WPI; 2001-226686/23.
XX
PT   Human 1p36 homozygosity deletion domain from the 36-position of first
PT   chromosome short arm in human neuroblastoma cell lines, applicable e.g.
PT   in gene diagnosis of tumors as well as in developing anti-cancer drugs.
XX
PS   Example 8; Page 93-95; 226pp; Japanese.
XX
CC   The present invention describes a homozygosity deletion domain co-
CC   existing in the 36-position of the first chromosome short arm (1p36) in
CC   human neuroblastoma. Also described are base sequences from the 1p36
CC   position of human neuroblastoma cell lines (NB-1 and MASS-NB-SCH-1),
CC   which are tumour suppressor genes in human neuroblastoma. The genes are
CC   tumour suppressor genes, base sequence data of which are applicable as
CC   tumour markers and reagents in studying mechanism of tumour body
CC   formation, and gene diagnosis of tumours as well as in developing anti-
CC   cancer drugs. AAF97787 to AAF97829 represent PCR primers used in the
CC   exemplification of the present invention, and AAF97830 to AAF97874
CC   represent sequences given in the exemplification of the present invention
XX
SQ   Sequence 6405 BP; 1650 A; 1449 C; 1383 G; 1923 T; 0 U; 0 Other;
```

```
  Query Match             16.7%;  Score 416.2;  DB 5;  Length 6405;
  Best Local Similarity   59.9%;  Pred. No. 7e-76;
  Matches  856;  Conservative    1;  Mismatches  504;  Indels   68;  Gaps    7;

Qy        603 GTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAG 662
Db       3636 GTCTTTATTTTTTATTTTTTTGAGGTGAGTTTCACTCTTGTTGCCCAGGCTGGAGTGCAA 3577

Qy        663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTC 722
Db       3576 TGGCACGATCTCAGCTTACAACAACCTCTGCCTCCTGAGTTCAAGTGATTCTCCTGCCTC 3517

Qy        723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db       3516 AGCCTCCCGAGTAGCTGGGATTACAGGCATGCACCACCACACTCAGCTGATTTTTTGTA 3457

Qy        783 TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db       3456 TTTTTAGTAGAGA-TGGGGTTCTCCATGTTGGTCAGGCTAGTCTCAAACTCCCGACCTCA 3398

Qy        843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db       3397 GGTGATCCGCCTGCCTCGGCCTCCCGAAGTGCTGGGATTACAGGCGTGAGCCATAAGGCC 3338

Qy        903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db       3337 TGGCCTATTTTTTACT---------------------------------------------- 3322

Qy        963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db       3321 -TTTAAGATGGAGTTTTGCTCTTGTTGCCCAAGCTGGAGTGCAATGGCATGATC------ 3269

Qy       1023 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 1082
Db       3268 -----TCGGCTCACTGCAACCTCCGCCTCCCAGGTTCAAGTGATTCTCTTGCCTCAGTCT 3214

Qy       1083 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTA 1142
Db       3213 CCCAAGTAGCTGGGATTACAGGCATGCACCACCATGCCCGGCTAA--ATTTGTATTTTTA 3156

Qy       1143 GCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGAT 1202
Db       3155 GTAGAGATGGGGTTTCTCCATGTTGGTTAGGCTGGTCTCAAACTCCCAACCTC-AGTGAT 3097

Qy       1203 CTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCG 1262
Db       3096 GCACCAGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGGCCG 3037
```

```
Qy       1263 ATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAG 1322
Db       3036 TACGTATCTTTATATCAGGATGTTTATAACAATTTTCTCTTAGAACATCTTTAGAATTTA 2977

Qy       1323 TAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTT 1382
Db       2976 ATACTAGGAAAAAAGGTTT-------CTCTCGACAAACTATTAGGCTTTCTTTTACCAGA 2924

Qy       1383 TTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACT 1442
Db       2923 GTGAAAATTGTACAATTTATCTTTCCCTTTTTGCATTAAATTCCATGTTCCAAATTAAAT 2864

Qy       1443 TATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCAT 1502
Db       2863 TACAATAACCACATTAGCCGAGGCAACCTATACATTTAGTTTTTCTTGTTACATAAATTT 2804

Qy       1503 TGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCA 1562
Db       2803 TTACTTATTTATATTTTAATAGCTTCACTGTATATATTTGTGGTAAGCTGCTCAAATCCT 2744

Qy       1563 CTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATA 1622
Db       2743 TTCTAGAAGTAGTAACTGGTATAAATTATAAATAAATAAAAACTTCCTGTAATAGCCTAT 2684

Qy       1623 CACACACACACACACACTC-ACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA 1681
Db       2683 TCTGCTATTTAACACTGTAAATATGTTGTGTAGTCAAAGTTCTAGTTCCTAAAAACAAAT 2624

Qy       1682 ACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGAC 1741
Db       2623 AAAACCAAACGACACAGCAAGCTTTTAAAAGTTACATAAAGATCGGCTGGGCATCGTGGC 2564

Qy       1742 TCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAA 1801
Db       2563 TCACGCGTATAATCCCAACACTTTGGGAGGCCAAGGCGGGTGGATCACTTGAGGTCAGGA 2504

Qy       1802 GTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTA 1861
Db       2503 GTTCAAGACCAGCCTGGCCAACATGATGAAACGCTGTCTCTACTAAAAATACAAAAATTA 2444

Qy       1862 GCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAAT 1921
Db       2443 GCTGGGTGTGATGGCACATGACTGTAATCCCAGCTACTCGGGAGGCTGAGGCAGGAGAAT 2384

Qy       1922 TGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCC 1981
Db       2383 CACTTGAACCCGGGAGATGAAGGTTGCAGTGAGCCAAGGTCACACCACTGCACTCTAGCC 2324

Qy       1982 TGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db       2323 TGGATGACAGAGCGAGACTCTGTCTCAAAAAAAAAAAAAGTTGCACAAA 2275
```

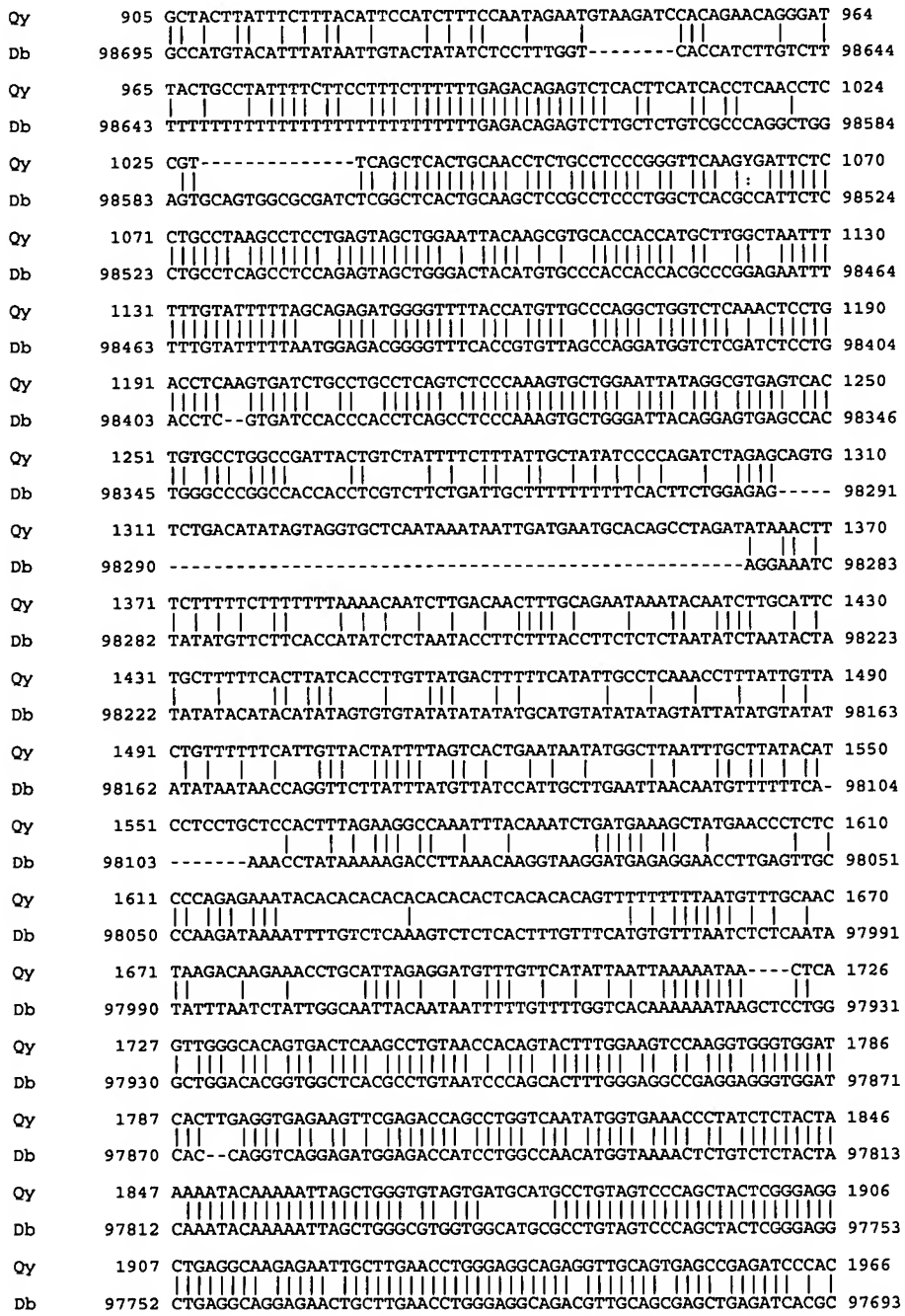
RESULT 11
ADP13332/c
ID ADP13332 standard; DNA; 126990 BP.
XX
AC ADP13332;
XX
DT 26-AUG-2004 (first entry)
XX
DE Renal cell carcinoma differentially expressed gene #68.
XX
KW ds; diagnosis; non-blood disease; solid tumor; gene expression;
KW peripheral blood mononuclear cell; renal cell carcinoma; prostate cancer;
KW head/neck cancer; differential expression.
XX
OS Homo sapiens.
XX
PN WO2004048933-A2.
XX
PD 10-JUN-2004.
XX
PF 21-NOV-2003; 2003WO-US037481.
XX
PR 21-NOV-2002; 2002US-0427982P.
PR 03-APR-2003; 2003US-0459782P.
XX
PA (AMHP) WYETH.
PA (TWIN/) TWINE N C.
PA (BURC/) BURCZYNSKI M E.
PA (TREP/) TREPICCHIO W L.
PA (DORN/) DORNER A.
PA (STOV/) STOVER J A.
PA (SLON/) SLONI D K.
XX
PI Twine NC, Burczynski ME, Trepicchio WL, Dorner A, Stover JA;
PI Sloni DK;
XX
DR WPI; 2004-460799/43.
XX
PT Diagnosing non-blood disease such as solid tumor, involves comparing
PT differential expression profile of specific genes in peripheral blood
PT sample of subject with reference expression profile of specific genes.
XX
PS Disclosure; SEQ ID NO 68; 350pp; English.
XX
CC The invention relate to a method of diagnosing (M1) non-blood disease
CC such as solid tumor by providing peripheral blood sample of human having
CC non-blood disease, and comparing an expression profile of specific genes
CC in the peripheral blood sample to reference expression profile of the
CC genes, where each of the genes is differentially expressed in peripheral
CC blood mononuclear cells (PBMCs) of patients having the disease as
CC compared to PBMCs of normal humans. The method is useful for diagnosing
CC non-blood disease such as solid tumor. The solid tumor is chosen from
CC renal cell carcinoma (RCC), prostate cancer and head/neck cancer. The
CC peripheral blood sample comprises enriched PBMCs. The peripheral blood
CC sample is a whole blood sample (claimed). (M1) is useful for identifying
CC genes that are differentially expressed in peripheral blood samples
CC isolated at different stages of progression, development or treatment of
CC RCC and/or other solid tumors. This sequence corresponds to a gene that
CC is differentially expressed and detected by the method of the invention.
CC (Note: this sequence is not given as part of the printed specification
CC but was obtained from WIPO in electronic format at
CC ftp.wipo./pub/published_pct_sequences).
XX
SQ Sequence 126990 BP; 36683 A; 27389 C; 27065 G; 35853 T; 0 U; 0 Other;

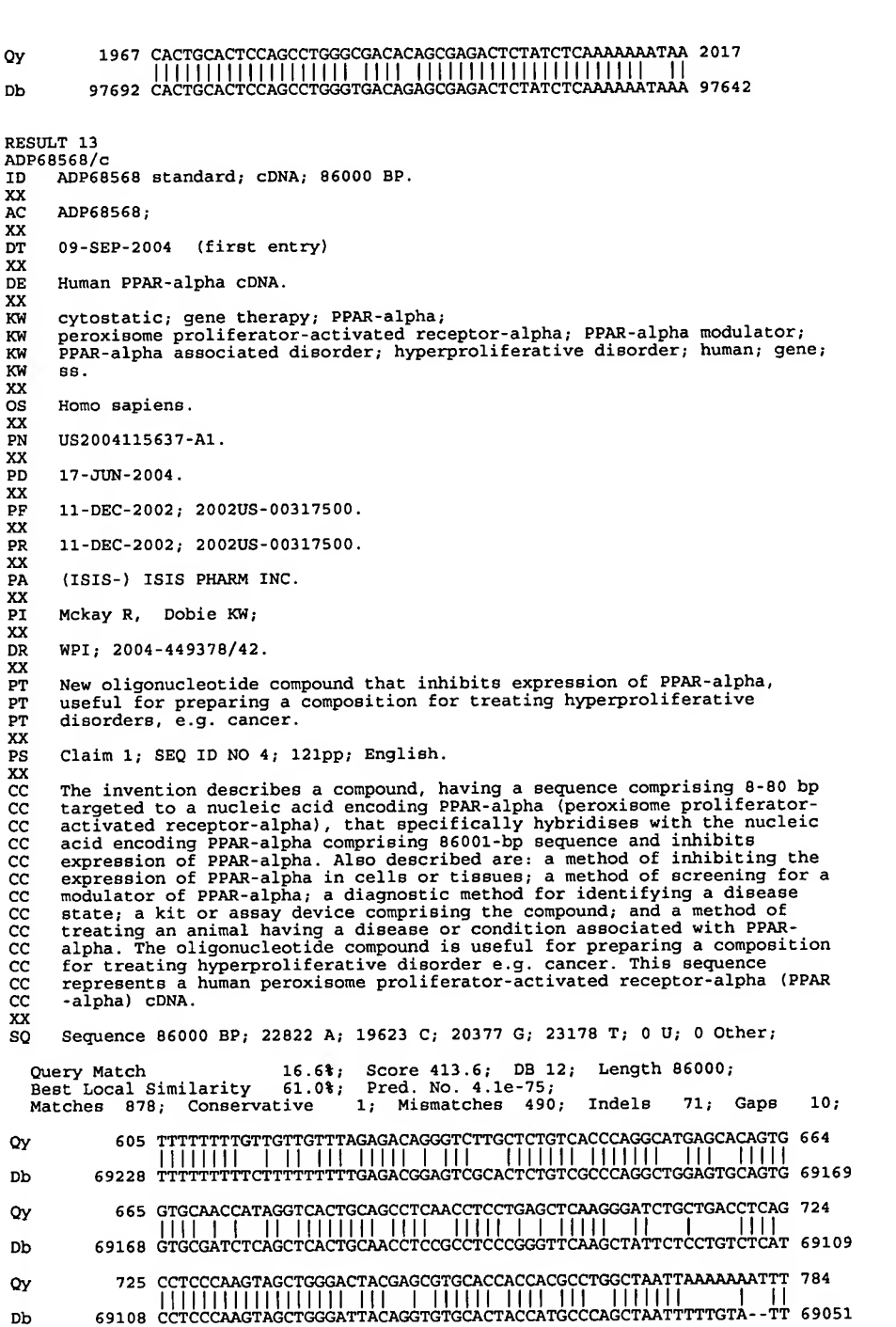
Query Match 16.6%; Score 416; DB 12; Length 126990;
Best Local Similarity 60.2%; Pred. No. 1.4e-75;
Matches 859; Conservative 1; Mismatches 536; Indels 32; Gaps 9;

```
Qy        611 TTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAA 670
Db      27332 TTTCTCTAATGCTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGTGCAGTGGTGTGA 27273

Qy        671 CCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCC 730
Db      27272 TCTCA-----------CCTCCACCTCCCAGGTTCAAGCGATTCTCCTGCCTCAGCCT-CC 27225

Qy        731 AAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGT 790
Db      27224 TGGTAGCTGGGAATACAGGCGTGCACCACCATGCCCGGCTAATTTTTATA--TTTTCAGT 27167

Qy        791 AGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCC 850
Db      27166 AGAGTCGCAGTTTCACCATGTTTCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCC 27107

Qy        851 TCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACT 910
Db      27106 ACCCGCCGCAGCCTCCCAACGTGCTGGGATAATAGGGGTGAGCCACTGCACCTGGCCTTT 27047

Qy        911 TATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGC 970
Db      27046 CTGTACTCTTTATTCTATTTCACTGGTCTGTTTCTTTTGGTTTCAGTATCATAGTACTTT 26987
```

```
Qy        971 CTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA 1030
              || | |  |||    | |  | | |   |   |||  |   |      | |    |||
Db      26986 TGTTTCTGTAGCTTAAAATAT--ATATATTTTAAGACTGGAATGCAGTGGCATGATCTCA 26929

Qy       1031 GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTA 1090
              ||||||||||||||||  |||||| || |||| : |||||  || || ||||||||||||
Db      26928 GCTCACTGCAACCTCCACCTCCCAGGCTCAAATTATTCTTGTGACTCAGCCTCCTGAGTA 26869

Qy       1091 GCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGAT 1150
              ||||| ||||||  ||  ||||||||||||    |||| ||||||||||||  || ||||
Db      26868 GCTGGGATTACAGGCAGGCACCACCATGCCCCACTAA-TTTTTGTATTTTCAGTAGAGTG 26810

Qy       1151 GGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGC 1210
              |  |||| |||||||| || |||||| ||| ||||||||||||| ||||||| ||| ||
Db      26809 GAAGTTTCACCATGTTTCCTAGGCTGATCTTAAACTCCTGACCTCGAGTGATCCGCCCGC 26750

Qy       1211 CTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGT 1270
              |||| ||||||||||||||| || | |||| | ||   ||| |    |||  |   | ||
Db      26749 GTCAGCCTCCCAAAGTGCTGGGATAACAGGCATAAGCTGCTGCGTGCAGCCTTTCTGTAT 26690

Qy       1271 CTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCT 1330
              || |  |  ||  |  |||      | | |      || | | |     |  ||  |
Db      26689 TCTTTATTCTGTTTCATTGATCTGTTCCTTTTGGTTTCAGTATCATACTATTTTTGTTTT 26630

Qy       1331 CAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTAAAAC 1390
              |  |||  | |  |||     |   |  | |  |  |     | ||  | | | |  |
Db      26629 GTTTCTATAGCTTAAAAATATGTATTTT---TTTTGAGATAGGGTCTCACTGTGTCACCC 26573

Qy       1391 AATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCT 1450
              |  || ||       ||  |      ||      |||| ||     |  |  |||  |
Db      26572 AGGCTGGAGTGTACTGGTGTGATCATGGCTCACTGCATCCTCAACCTCCTATACTCAA-T 26514

Qy       1451 TGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTA 1510
              || |    |   |||   | ||    | ||  |     |   ||   | ||| |||
Db      26513 TGATCCTCCCACCTCAGCCTCCCAAGTAGCTGCACCACCATGTTTGGCTAATTTTTAAAT 26454

Qy       1511 TTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAA 1570
              ||||  ||       | | | | ||     |   |          ||    ||||      |||
Db      26453 TTTTTGTAGAAACAGAGTCTTGCCATGTCACCCAGGCTGGTCTTGAACTCCTGGGCTGAA 26394

Qy       1571 GGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACA 1630
              |   | ||   |  | || ||     | | |  |       |||     ||    |
Db      26393 G---AGATCCGCTGACCTCATCCTCCCAAAGTGCTGGAATTACAGGTGTGCACCACCTTG 26337

Qy       1631 CACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAA--------A 1682
              | ||  ||   | |      ||      | |  |  |      ||   |  |
Db      26336 CCCAGCCAAAAATATGTTTTTGGCAGATCACCTGAGGTCGGGAGTTCAAGACCAGCCTGG 26277

Qy       1683 CCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACT 1742
              ||  |||   |  |  ||   ||   | ||  | | | || | |  ||||| ||||| ||
Db      26276 CCAACATGGCAAAACCTTGTCTCTACTAAAGAATACAAAAATTATCTGGGCGCAGTGGCT 26217

Qy       1743 CAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAG 1802
              || | ||||||   |||||| |||||| ||||||  |||||||||| |||||| || ||
Db      26216 CATGTCTGTAAATCCAGTACTCTGGAAGGCCAAGGCAGGTGGATCACCTGAGGTCAGGAG 26157

Qy       1803 TTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAG 1862
              ||  |||||||||||| ||| |||||||||||| |||||||||||||||||||||||||
Db      26156 TTTAAGACCAGCCTGGCCAACCTGGTGAAACCCTGTCTCTACTAAAAATACAAAAATTAG 26097

Qy       1863 CTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATT 1922
              ||||| || ||||  |||||||| |||||||||||| |||||||||||| ||||||
Db      26096 CTGGGCGTGGTGACCGGTGCCTGTAATCCCAGCTACTCAGGAGGCTGAGGCAGGAGAATC 26037

Qy       1923 GCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCT 1982
              |||||||||||||||| |||||||||||||||||||||||||    ||| |||||| |||||
Db      26036 ACTTGAACCTGGGAGGTAGAGGTTGCAGTGAGCCGAGATCATGCCATTGCACTCTAGCCT 25977

Qy       1983 GGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
              ||| |||| ||||||| ||| ||||||||||  ||| ||  |||| ||
Db      25976 GGGTGACAGAGCGAGATTCTGTCTCAAAAAAGAAAAAAAGAAAAAAAA 25929
```

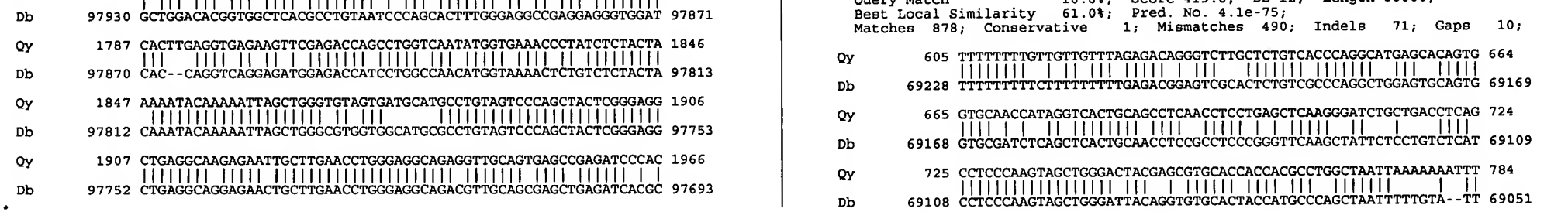
```
RESULT 12
ACN44170/c
ID   ACN44170 standard; DNA; 196686 BP.
XX
AC   ACN44170;
XX
DT   18-NOV-2004  (first entry)
XX
DE   Human genomic sequence hCG39530.
XX
KW   Cytostatic; carcinoma; lymphoma; cancer; human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2003073826-A2.
XX
PD   12-SEP-2003.
XX
PF   28-FEB-2003; 2003WO-US006235.
XX
PR   01-MAR-2002; 2002US-00087192.
XX
PA   (SAGR-) SAGRES DISCOVERY.
XX
PI   Morris DW;
XX
DR   WPI; 2003-328604/31.
XX
PT   Recombinant nucleic acid useful for diagnosis and treatment of carcinoma
PT   comprises a nucleotide sequence.
XX
PS   Claim 1; SEQ ID NO 484; 0pp; English.
XX
CC   The present invention relates to novel DNA and protein sequences which
CC   are associated with carcinomas. The sequences are useful for: (i) for
CC   screening drug candidates; (ii) for screening of bioactive agent capable
CC   of binding to Carcinoma Associated Protein (CAP); (iii) for screening of
CC   a bioactive agent capable of modulating the activity of CAP; (iv) for
CC   evaluating the effect of a candidate carcinoma drug; (v) for diagnosing
CC   carcinoma; (vi) for inhibiting the activity of CAP; (vii) for treating
CC   carcinoma; (viii) for neutralizing the effect of CAP; (ix) as a biochip;
CC   (x) for diagnosing carcinoma or a propensity to carcinoma; and (xi) for
CC   determining Carcinoma Associated (CA) gene copy number. In addition, the
CC   CA genes are useful as DNA vaccines and the CAP are useful as markers of
CC   carcinoma including lymphoma. The present sequence is one such CA coding
CC   sequence. Note: This patent is an equivalent to basic patent
CC   US2002182586A1, for which no sequence data was published
XX
SQ   Sequence 196686 BP; 53978 A; 42758 C; 43862 G; 55372 T; 0 U; 716 Other;

  Query Match             16.6%;  Score 414.2;  DB 11;  Length 196686;
  Best Local Similarity   60.7%;  Pred. No. 3.7e-75;
  Matches  869;  Conservative    1;  Mismatches  464;  Indels   97;  Gaps    8;

Qy        605 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTG 664
              ||  ||   || ||  || ||||  ||||||||||||||    ||||||  |||   | ||
Db      98993 TTCATTCATTTATTCATTCGAGATGGGGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAATG 98934

Qy        665 GTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAG 724
              |||||| |   | |||||||| |||| |||||| |   |||||  |||   |   ||||||
Db      98933 GTGCAATCTCGGCTCACTGCAACCTCCACCTCCAGGATTCAAGTGATTCTCCTGCCTCAG 98874

Qy        725 CCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTT 784
              |||||  |||| ||||||  |||   |      |||||||  || |||||||      |  ||
Db      98873 CCTCCTGAGTAACTGGGATTACAGACACCTGCCACCACACCCGGCTAATTTTTGTA--TT 98816

Qy        785 TTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAG 844
              ||| |||  |||  | || | || | |||||||||  || ||||  |||||||||  | ||||
Db      98815 TTTAGTAAAGATGGAGTTTCACCATGTTGGCCAGCCTGGTCTCAAACTCCTGACCTCAAA 98756

Qy        845 CAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCG 904
              |  |||| || |||||||| |||||||||||||||||||||||||   ||||||||   | |
Db      98755 GGATCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCA 98696
```

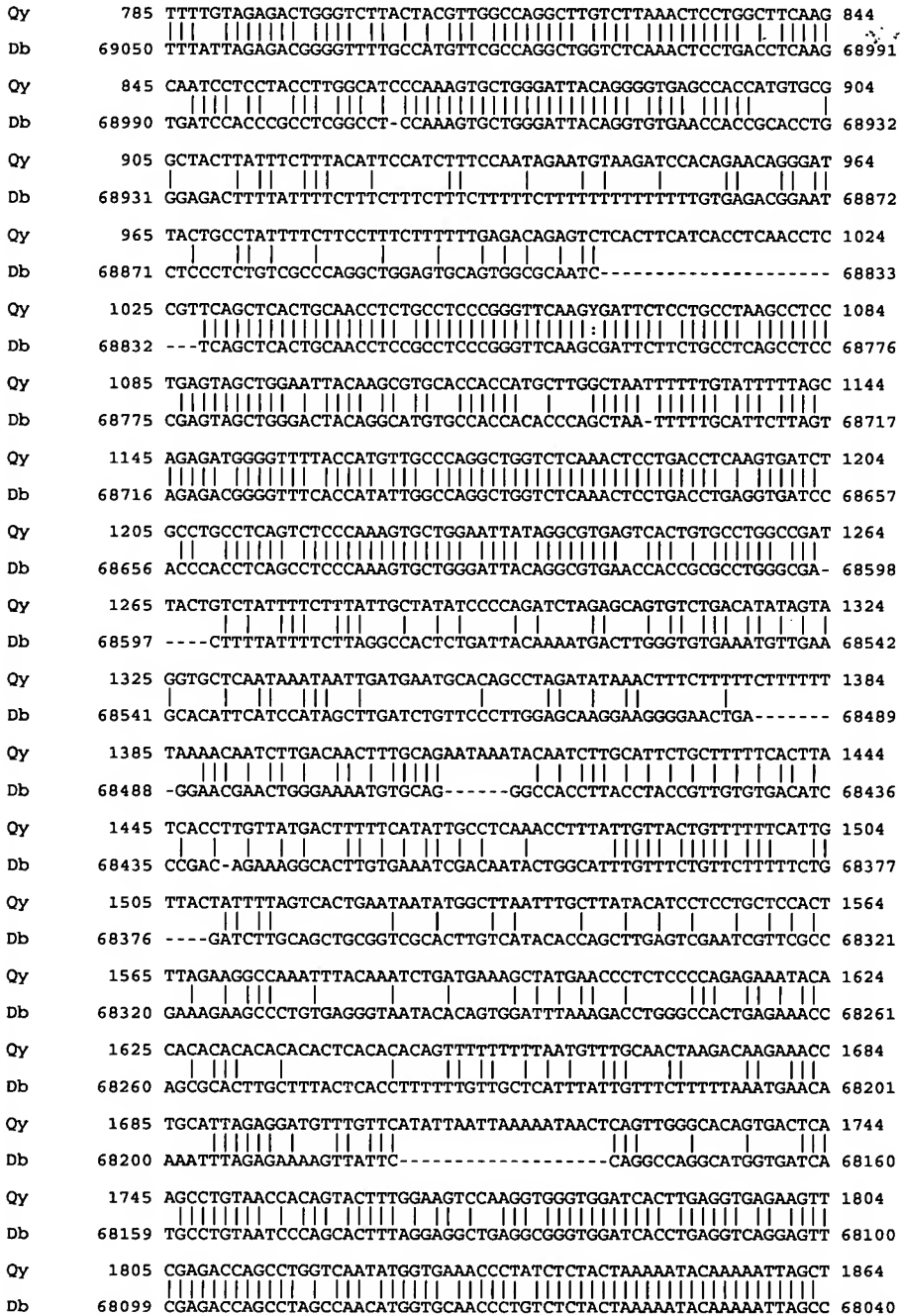


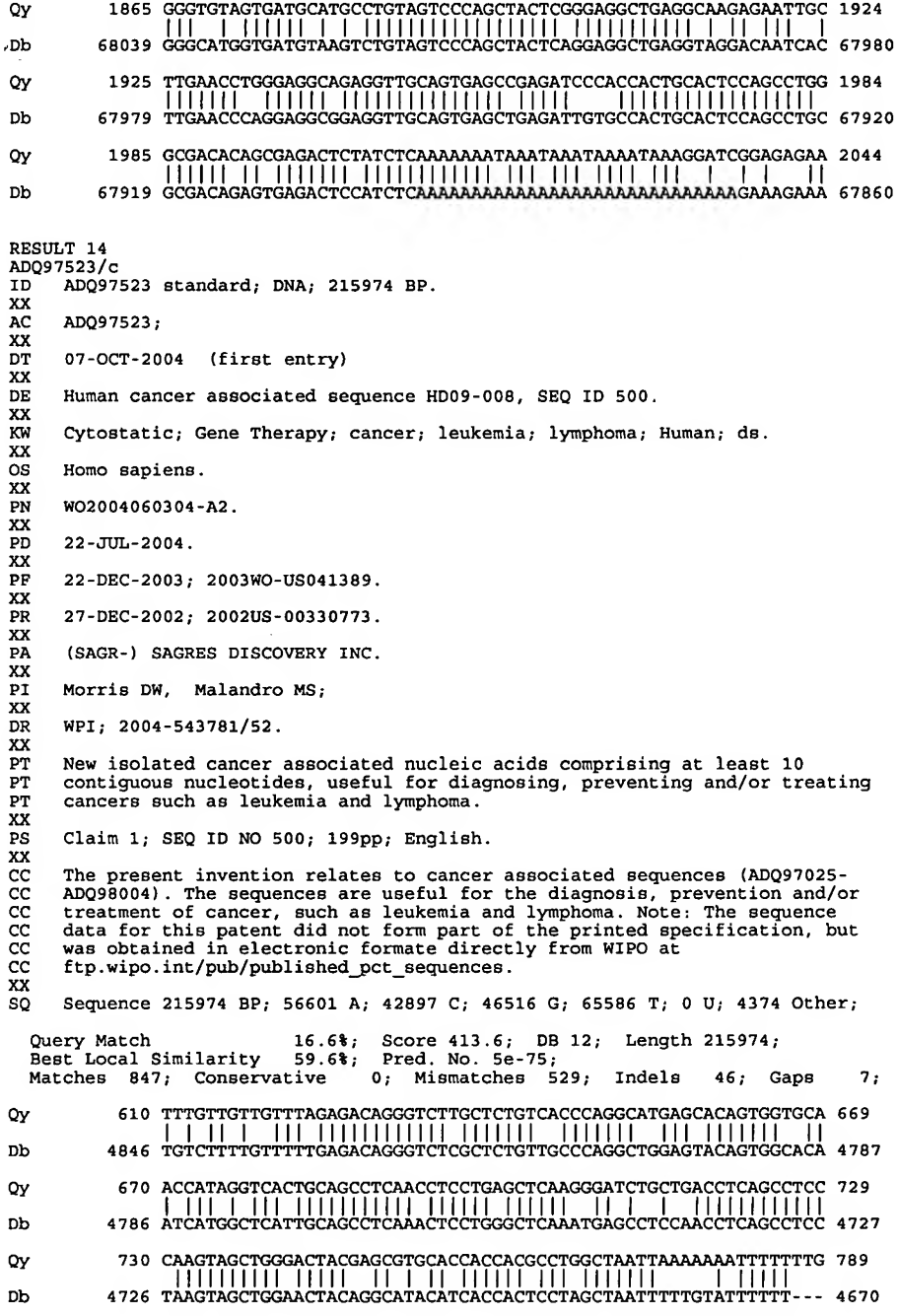


RESULT 13
ADP68568/c
ID ADP68568 standard; cDNA; 86000 BP.
XX
AC ADP68568;
XX
DT 09-SEP-2004 (first entry)
XX
DE Human PPAR-alpha cDNA.
XX
KW cytostatic; gene therapy; PPAR-alpha;
KW peroxisome proliferator-activated receptor-alpha; PPAR-alpha modulator;
KW PPAR-alpha associated disorder; hyperproliferative disorder; human; gene;
KW ss.
XX
OS Homo sapiens.
XX
PN US2004115637-A1.
XX
PD 17-JUN-2004.
XX
PF 11-DEC-2002; 2002US-00317500.
XX
PR 11-DEC-2002; 2002US-00317500.
XX
PA (ISIS-) ISIS PHARM INC.
XX
PI Mckay R, Dobie KW;
XX
DR WPI; 2004-449378/42.
XX
PT New oligonucleotide compound that inhibits expression of PPAR-alpha,
PT useful for preparing a composition for treating hyperproliferative
PT disorders, e.g. cancer.
XX
PS Claim 1; SEQ ID NO 4; 121pp; English.
XX
CC The invention describes a compound, having a sequence comprising 8-80 bp
CC targeted to a nucleic acid encoding PPAR-alpha (peroxisome proliferator-
CC activated receptor-alpha), that specifically hybridises with the nucleic
CC acid encoding PPAR-alpha comprising 86001-bp sequence and inhibits
CC expression of PPAR-alpha. Also described are: a method of inhibiting the
CC expression of PPAR-alpha in cells or tissues; a method of screening for a
CC modulator of PPAR-alpha; a diagnostic method for identifying a disease
CC state; a kit or assay device comprising the compound; and a method of
CC treating an animal having a disease or condition associated with PPAR-
CC alpha. The oligonucleotide compound is useful for preparing a composition
CC for treating hyperproliferative disorder e.g. cancer. This sequence
CC represents a human peroxisome proliferator-activated receptor-alpha (PPAR
CC -alpha) cDNA.
XX
SQ Sequence 86000 BP; 22822 A; 19623 C; 20377 G; 23178 T; 0 U; 0 Other;

Query Match 16.6%; Score 413.6; DB 12; Length 86000;
Best Local Similarity 61.0%; Pred. No. 4.1e-75;
Matches 878; Conservative 1; Mismatches 490; Indels 71; Gaps 10;

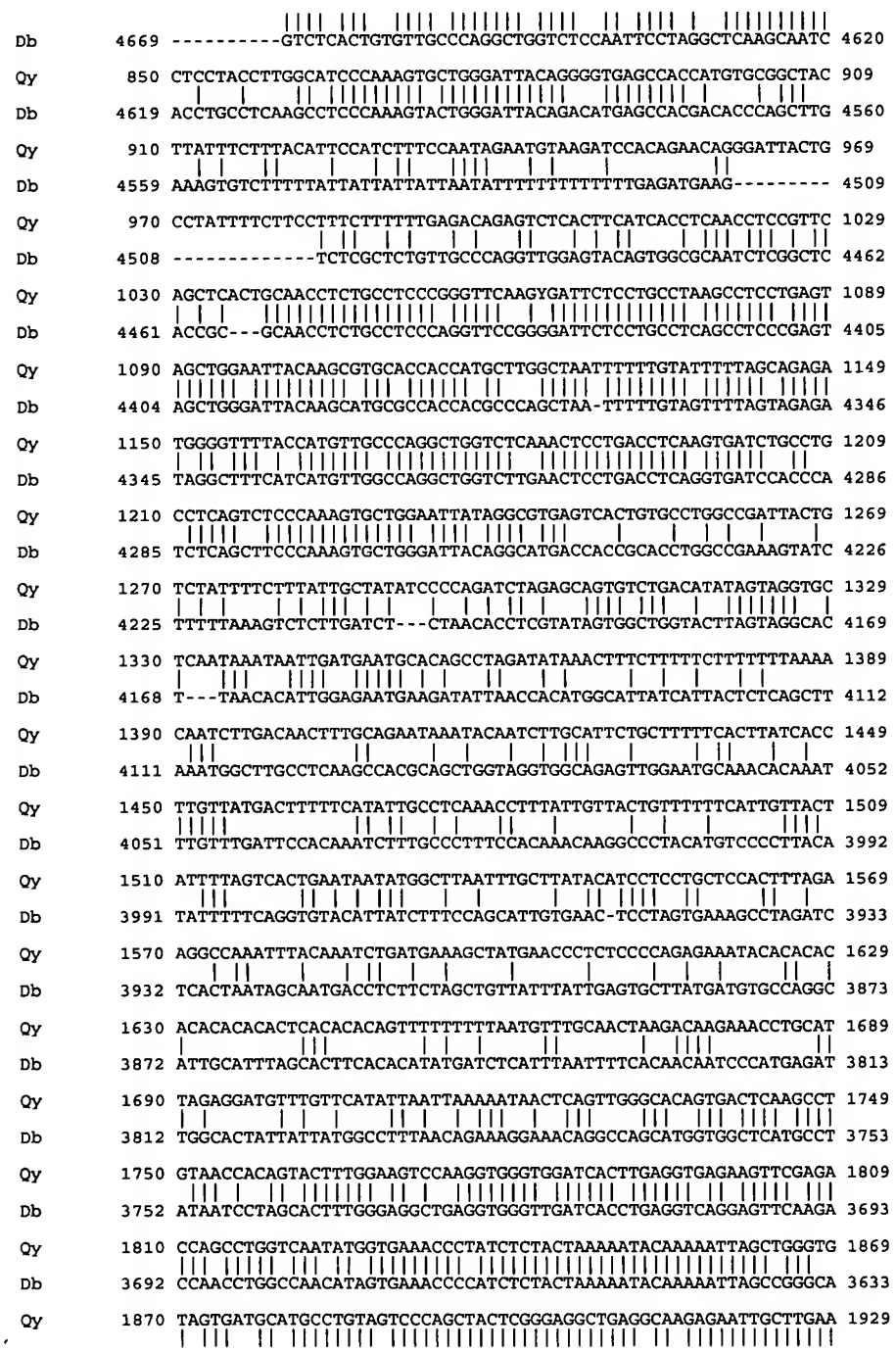


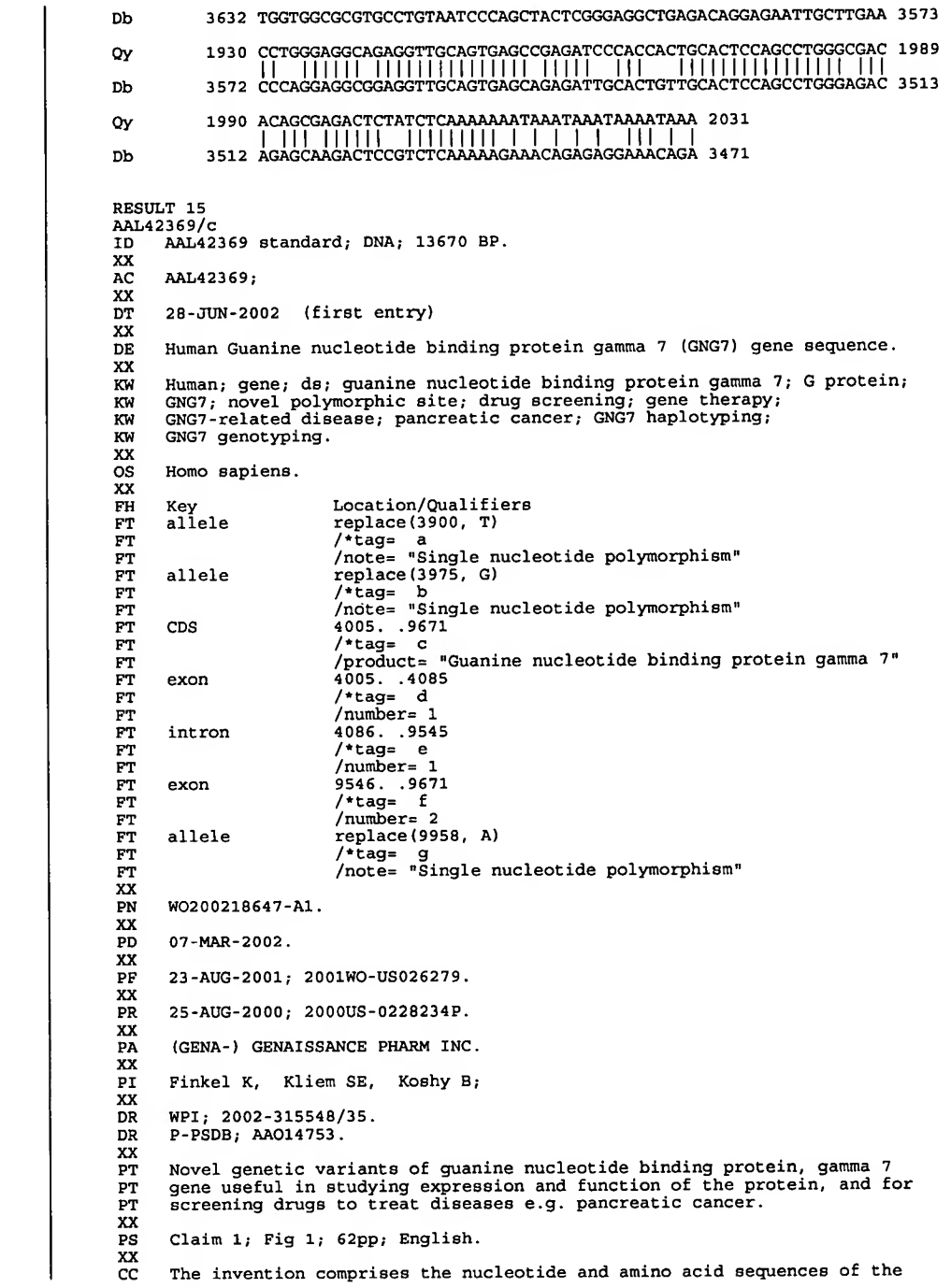




RESULT 14
ADQ97523/c
ID ADQ97523 standard; DNA; 215974 BP.
XX
AC ADQ97523;
XX
DT 07-OCT-2004 (first entry)
XX
DE Human cancer associated sequence HD09-008, SEQ ID 500.
XX
KW Cytostatic; Gene Therapy; cancer; leukemia; lymphoma; Human; ds.
XX
OS Homo sapiens.
XX
PN WO2004060304-A2.
XX
PD 22-JUL-2004.
XX
PF 22-DEC-2003; 2003WO-US041389.
XX
PR 27-DEC-2002; 2002US-00330773.
XX
PA (SAGR-) SAGRES DISCOVERY INC.
XX
PI Morris DW, Malandro MS;
XX
DR WPI; 2004-543781/52.
XX
PT New isolated cancer associated nucleic acids comprising at least 10
PT contiguous nucleotides, useful for diagnosing, preventing and/or treating
PT cancers such as leukemia and lymphoma.
XX
PS Claim 1; SEQ ID NO 500; 199pp; English.
XX
CC The present invention relates to cancer associated sequences (ADQ97025-
CC ADQ98004). The sequences are useful for the diagnosis, prevention and/or
CC treatment of cancer, such as leukemia and lymphoma. Note: The sequence
CC data for this patent did not form part of the printed specification, but
CC was obtained in electronic formate directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 215974 BP; 56601 A; 42897 C; 46516 G; 65586 T; 0 U; 4374 Other;

Query Match 16.6%; Score 413.6; DB 12; Length 215974;
Best Local Similarity 59.6%; Pred. No. 5e-75;
Matches 847; Conservative 0; Mismatches 529; Indels 46; Gaps 7;





RESULT 15
AAL42369/c
ID AAL42369 standard; DNA; 13670 BP.
XX
AC AAL42369;
XX
DT 28-JUN-2002 (first entry)
XX
DE Human Guanine nucleotide binding protein gamma 7 (GNG7) gene sequence.
XX
KW Human; gene; ds; guanine nucleotide binding protein gamma 7; G protein;
KW GNG7; novel polymorphic site; drug screening; gene therapy;
KW GNG7-related disease; pancreatic cancer; GNG7 haplotyping;
KW GNG7 genotyping.
XX
OS Homo sapiens.
XX

FH	Key	Location/Qualifiers
FT	allele	replace(3900, T)
FT		/*tag= a
FT		/note= "Single nucleotide polymorphism"
FT	allele	replace(3975, G)
FT		/*tag= b
FT		/note= "Single nucleotide polymorphism"
FT	CDS	4005. .9671
FT		/*tag= c
FT		/product= "Guanine nucleotide binding protein gamma 7"
FT	exon	4005. .4085
FT		/*tag= d
FT		/number= 1
FT	intron	4086. .9545
FT		/*tag= e
FT		/number= 1
FT	exon	9546. .9671
FT		/*tag= f
FT		/number= 2
FT	allele	replace(9958, A)
FT		/*tag= g
FT		/note= "Single nucleotide polymorphism"

XX
PN WO200218647-A1.
XX
PD 07-MAR-2002.
XX
PF 23-AUG-2001; 2001WO-US026279.
XX
PR 25-AUG-2000; 2000US-0228234P.
XX
PA (GENA-) GENAISSANCE PHARM INC.
XX
PI Finkel K, Kliem SE, Koshy B;
XX
DR WPI; 2002-315548/35.
DR P-PSDB; AAO14753.
XX
PT Novel genetic variants of guanine nucleotide binding protein, gamma 7
PT gene useful in studying expression and function of the protein, and for
PT screening drugs to treat diseases e.g. pancreatic cancer.
XX
PS Claim 1; Fig 1; 62pp; English.
XX
CC The invention comprises the nucleotide and amino acid sequences of the

```
CC   human guanine nucleotide binding protein (G protein) gamma 7 (GNG7). The
CC   invention specifically relates to the discovery of three novel
CC   polymorphic sites in the GNG7 gene. The GNG7 nucleotide and protein
CC   sequences are useful for screening for drugs which target GNG7 and may be
CC   used to treat (gene therapy) GNG7-related diseases (e.g. pancreatic
CC   cancer). The GNG7 nucleotide sequence can be used for haplotyping GNG7,
CC   genotyping GNG7 and predicting the haplotype pair for GNG7 in an
CC   individual. The present DNA sequence represents the human GNG7 gene
CC   sequence
XX
SQ   Sequence 13670 BP; 3047 A; 3946 C; 3680 G; 2997 T; 0 U; 0 Other;

  Query Match             16.5%;  Score 413.4;  DB 6;  Length 13670;
  Best Local Similarity   60.6%;  Pred. No. 3.1e-75;
  Matches  876;  Conservative    1;  Mismatches  512;  Indels   56;  Gaps   10;

Qy        604 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCAT-GAGCACAG 662
Db       5900 TTTGTTTTTGTTTTTTTCTGAGATGGAGTCTCACCCTGTCGCTTGGGGTTGGAGGGCAG 5841

Qy        663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGATCTGCTGACCTC 722
Db       5840 TGGCGCAATCTCAGCTCACTGCAACCTCTGCCTCTCGGGTTCAAGCTATTCTCCTGACTC 5781

Qy        723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db       5780 AGTCTCCCAAGTAGCTGGGACTATAGGAGCCTGCCACCACGCCTGGCTAATTTTTTTGTA 5721

Qy        783 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db       5720 ATTTTAGTAGAGATGGGGTTTCACCATATTGGCCAGGCTGGTCTCGAACTCCTGAC--CT 5663

Qy        843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db       5662 TGTGATCCGCCCACCTCAGCCTCCCAAAGTGCTGCGATTCCAGGCGTGAGTCACCGTGCC 5603

Qy        903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db       5602 TGGCCAGTTAAATGTTTTTTTAGATGTTGCTGTGAAGGTTGTTTTTTTC----------- 5554

Qy        963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db       5553 -TTGTTTCTTTTTTTTTTTTTTTTTTTTTTGAGATAGAGTCTCGCTCTGTTCCCCAGGCT 5495

Qy       1023 TCCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 1068
Db       5494 GGAGTGCAATGGTGCGATCTCGGCTCACTGCAACCTCCACCTCCCAGGTTCAAGTGATTC 5435

Qy       1069 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA- 1127
Db       5434 TCCTGCCTCAGCCTCCCAAATAGCTGGGATTACAGGCATGCACCACCATACCCGGGTAAT 5375

Qy       1128 ---TTTTTGTATTTTTAGCAGAGATGGGGTTTACCATGTTGCCCAGGCTGGTCTCAAA 1184
Db       5374 CTTTTTTTGTATTTTTAGGAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAA 5315

Qy       1185 CTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTG 1244
Db       5314 CCCCTGACCTCAGGTGATCTGCCCACCTCAGCTTCCCAAAGTGCTGGAATTATAAGTGTG 5255

Qy       1245 AGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGA 1304
Db       5254 AGTCACCGCGCTCAGCCTGTT-TTTTATTTTGTAGAGATGGGGTCTTGCTATGTTGTTGC 5196

Qy       1305 GCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATAT 1364
Db       5195 CCAGGCTGGAATTGAACGCCTGACCTCAAGTGATCCTCCTGCCTCGGCCTCCCAATGTGC 5136

Qy       1365 AAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTT 1424
Db       5135 TGGGATT----------------ACAGCTGTGAGCCACTGTGCCTCACATGCAAAATTTT 5092

Qy       1425 GCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1484
Db       5091 TTTTTTTTTTTTGATGTGATTCACATT----TAAATCCGTAGACTTCAAGCACAGCAGGT 5036

Qy       1485 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1544
Db       5035 GACCGTCTGTAATGCAGATGGGCCTCATCTAATCAGTTGAAGGGCTTAAGAGAAAAAGCC 4976

Qy       1545 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1604
Db       4975 TGGGGTCCTTCCAGGAGGAGGGAACTCTGCCTCCAGGTGGCCTTGGGGTTCCAGCTGCAG 4916

Qy       1605 CCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTAATGTT 1664
Db       4915 CACCAACTCTTCCCTGGGTCTCCAGCCTGCTGGCCTACACTGCAGATTTCAGAGCTTGCC 4856

Qy       1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724
Db       4855 TTAAAATAAATCTCGACAA-GGCGCAGCCTCAAGTGTGATCCTTGTAAAAAACTTCAAAC 4797

Qy       1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784
Db       4796 GGGCCAGGCATGGTGGCTCACGCCTGGAATCCCAGGACTTTGGGAGGCCGAGGCAGGCGG 4737

Qy       1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844
Db       4736 ATCACCTGAAGTTAGGAGTTCAAGACCAGCCTGGGCAACATGGCGAAACGCCGTCTCTAC 4677

Qy       1845 TAAAAATAC-AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGG 1903
Db       4676 TAAAAATACAAAAAAATTAGCCAGGCATGGTGGTGCATGCCTGTAATCCCAGCTATTGGG 4617

Qy       1904 AGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCC 1963
Db       4616 AGGTTGAGGCAGAAGAATTGCTTGAACCCGGGAGGCAGAGGTTGCAGTGAGCCGAGATCG 4557

Qy       1964 CACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAAT 2023
Db       4556 CGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCCGTCTCAAACAAACAAAACAAA 4497

Qy       2024 AAAAT 2028
Db       4496 CAAAT 4492

Search completed: February 17, 2006, 10:20:49
Job time : 1487 secs
```

OM nucleic - nucleic search, using sw model

Run on: February 17, 2006, 09:56:20 ; Search time 12639 Seconds
(without alignments)
11248.158 Million cell updates/sec

Title: US-10-607-806-1-C7256_COPY_7000_9500
Perfect score: 2499
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_in:*
3: gb_env:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pr:*
9: gb_ro:*
10: gb_sts:*
11: gb_sy:*
12: gb_un:*
13: gb_vi:*
14: gb_htg:*
15: gb_pl:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

	Result No.	Score	% Query Match	Length	DB	ID	Description
	1	2497.4	99.9	8368	8	AY438977	AY438977 Homo sapi
	2	2497.4	99.9	13612	6	AX377239	AX377239 Sequence
c	3	2497.4	99.9	122302	8	AC003982	AC003982 Homo sapi
c	4	2497.4	99.9	220384	14	AC078926	AC078926 Homo sapi
	5	2111.2	84.5	189729	14	AC073930	AC073930 Homo sapi
c	6	462.6	18.5	178887	14	AC068981	AC068981 Homo sapi
	7	462.6	18.5	190162	8	AC009244	AC009244 Homo sapi
	8	460.6	18.4	123192	14	AP000589	AP000589 Homo sapi
	9	460.4	18.4	150266	8	AC022408	AC022408 Homo sapi
	10	459.4	18.4	167996	8	AC021753	AC021753 Homo sapi
	11	457.4	18.3	137693	8	AC003689	AC003689 Homo sapi
	12	456.2	18.3	201460	8	AP003721	AP003721 Homo sapi
	13	451.2	18.1	195986	14	AC116933	AC116933 Papio anu
c	14	449.2	18.0	134210	8	AC005052	AC005052 Homo sapi
c	15	447.8	17.9	131943	8	AC005484	AC005484 Homo sapi
c	16	447.8	17.9	136119	8	HS1028D15	AL121886 Human DNA
	17	447.8	17.9	186197	14	AL627384	AL627384 Homo sapi
	18	446.2	17.9	163681	8	AL136992	AL136992 Human DNA
	19	445	17.8	117336	8	AC093171	AC093171 Homo sapi
	20	442.6	17.7	91927	8	AC004771	AC004771 Homo sapi
c	21	441.8	17.7	200430	8	AC011500	AC011500 Homo sapi
c	22	439.4	17.6	180076	14	AC022177	AC022177 Homo sapi
c	23	438.8	17.6	201227	14	AC160565	AC160565 Pan trogl
c	24	438.6	17.6	163704	14	AC141415	AC141415 Pan trogl
c	25	438.2	17.5	199045	14	AC149081	AC149081 Pan trogl
c	26	437.8	17.5	185574	14	AC161822	AC161822 Pan trogl
c	27	435.8	17.4	203790	8	AC010422	AC010422 Homo sapi
	28	435	17.4	119043	8	AC139451	AC139451 Homo sapi
c	29	435	17.4	163542	14	AC129071	AC129071 Pan trogl
c	30	434.8	17.4	96995	8	AC008149	AC008149 Homo sapi
	31	434.2	17.4	152659	8	AL591503	AL591503 Human DNA
c	32	433.8	17.4	199384	8	AP003531	AP003531 Homo sapi
	33	433.2	17.3	89448	8	AC005067	AC005067 Homo sapi
c	34	433.2	17.3	178976	14	AL391095	AL391095 Homo sapi
	35	432.6	17.3	156933	8	AC020750	AC020750 Homo sapi
c	36	432.2	17.3	81117	8	AC016898	AC016898 Homo sapi
c	37	432.2	17.3	192961	8	AP002770	AP002770 Homo sapi
c	38	432	17.3	153392	8	AL133548	AL133548 Human DNA
	39	432	17.3	235183	14	AC160951	AC160951 Colobus g
c	40	431.4	17.3	214031	14	AC160952	AC160952 Colobus g
c	41	431	17.2	134957	8	HSJ930L11	AL049759 Human DNA
c	42	430.8	17.2	44062	8	AC004517	AC004517 Homo sapi
c	43	430.8	17.2	118396	8	AC073138	AC073138 Homo sapi
	44	430.8	17.2	162701	14	AC073317	AC073317 Homo sapi
c	45	430.6	17.2	41613	8	AC004790	AC004790 Homo sapi

ALIGNMENTS

RESULT 1
AY438977
LOCUS AY438977 8368 bp DNA linear PRI 29-OCT-2003
DEFINITION Homo sapiens phospholipase A2, group IB (pancreas) (PLA2G1B) gene, complete cds.
ACCESSION AY438977
VERSION AY438977.1 GI:37953284
KEYWORDS .
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 8368)
AUTHORS Rieder,M.J., Livingston,R.J., Daniels,M.R., Chung,M.-W., Miyamoto,K.E., Nguyen,C.P., Nguyen,D.A., Poel,C.L., Robertson,P.D., Schackwitz,W.S., Sherwood,J.K., Witrak,L.A. and Nickerson,D.A.
TITLE Direct Submission
JOURNAL Submitted (16-OCT-2003) Genome Sciences, University of Washington, 1705 NE Pacific, Seattle, WA 98195, USA
COMMENT To cite this work please use: NIEHS-SNPs, Environmental Genome Project, NIEHS ES15478, Department of Genome Sciences, Seattle, WA (URL: http://egp.gs.washington.edu).
FEATURES Location/Qualifiers
source 1. .8368
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 101. .257
/rpt_family="MIR"
/rpt_type=dispersed
variation 205
/frequency="0.01"
/replace="a"
variation 306
/frequency="0.01"
/replace="a"
repeat_region 326. .538
/rpt_family="Alu"
/rpt_type=dispersed
misc_feature 354. .496

```
                /note="Region not scanned for variation"
gene            875. .6548
                /gene="PLA2G1B"
mRNA            join(875. .944,2644. .2803,3603. .3730,6347. .6548)
                /gene="PLA2G1B"
                /product="phospholipase A2, group IB (pancreas)"
CDS             join(911. .944,2644. .2803,3603. .3730,6347. .6471)
                /gene="PLA2G1B"
                /codon_start=1
                /product="phospholipase A2, group IB (pancreas)"
                /protein_id="AAR05441.1"
                /db_xref="GI:37953285"
                /translation="MKLLVLAVLLTVAAADSGISPRAVWQFRKMIKCVIPGSDPFLEY
                NNYGCYCGLGGSGTPVDELDKCCQTHDNCYDQAKKLDSCKFLLDNPYTHTYSYSCSGS
                AITCSSKNKECEAFICNCDRNAAICFSKAPYNKAHKNLDTKKYCQS"
repeat_region   1174. .1307
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1286
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="g"
repeat_region   1319. .1588
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       1343
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="a"
variation       1593
                /gene="PLA2G1B"
                /frequency="0.02"
                /replace="t"
repeat_region   1605. .1669
                /rpt_family="L2"
                /rpt_type=dispersed
misc_feature    1638. .2188
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
repeat_region   1670. .1975
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   1976. .2038
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2086. .2165
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   2194. .2319
                /rpt_family="Alu"
                /rpt_type=dispersed
misc_feature    2319. .2453
                /gene="PLA2G1B"
                /note="Region not scanned for variation"
variation       3027
                /gene="PLA2G1B"
                /frequency="0.25"
                /replace="t"
repeat_region   3165. .3431
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       3262
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3321
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
variation       3396
                /gene="PLA2G1B"
                /frequency="0.03"
                /replace="g"
variation       3630
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="c"
variation       3702
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="a"
variation       3768
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       3818
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="a"
repeat_region   3846. .3964
                /rpt_family="MIR"
                /rpt_type=dispersed
repeat_region   3977. .4280
                /rpt_family="Alu"
                /rpt_type=dispersed
repeat_region   4281. .4346
                /rpt_family="L2"
                /rpt_type=dispersed
repeat_region   4347. .4634
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       4386
                /gene="PLA2G1B"
                /frequency="0.08"
                /replace="t"
repeat_region   4635. .4723
                /rpt_family="L2"
                /rpt_type=dispersed
variation       5013
                /gene="PLA2G1B"
                /frequency="0.10"
                /replace="a"
repeat_region   5103. .5404
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5133
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5543
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="c"
variation       5555
                /gene="PLA2G1B"
                /frequency="0.01"
                /replace="t"
variation       5556
                /gene="PLA2G1B"
                /frequency="0.15"
                /replace="g"
repeat_region   5577. .5629
                /rpt_family="L1"
                /rpt_type=dispersed
variation       5580
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
variation       5675
                /gene="PLA2G1B"
                /frequency="0.09"
                /replace="g"
repeat_region   5779. .6083
                /rpt_family="Alu"
                /rpt_type=dispersed
variation       5929
```

```
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       5949
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="g"
     variation       6039
                     /gene="PLA2G1B"
                     /frequency="0.01"
                     /replace="a"
     variation       6082
                     /gene="PLA2G1B"
                     /frequency="0.14"
                     /replace="t"

  Query Match             99.9%;  Score 2497.4;  DB 8;  Length 8368;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2495;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3374 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 3433

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3434 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 3493

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3494 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 3553

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3554 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 3613

Qy        241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       3614 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 3673

Qy        301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::|||||||||||||||||||||||:||||||||||||||||||||||||||||||||||
Db       3674 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 3733

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3734 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 3793

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3794 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 3853

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3854 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 3913

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3914 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 3973

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       3974 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 4033

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4034 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 4093

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4094 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 4153

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4154 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 4213

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4214 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 4273

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4274 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 4333

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4334 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 4393

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db       4394 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 4453

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4454 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 4513

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4514 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 4573

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4574 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 4633

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4634 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 4693

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4694 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 4753

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4754 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 4813

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4814 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 4873

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4874 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 4933

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4934 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 4993

Qy       1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       4994 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 5053

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5054 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 5113

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5114 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 5173

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5174 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 5233

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db       5234 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 5293

Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5294 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 5353

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5354 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 5413

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5414 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 5473

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5474 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 5533

Qy       2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
             |||||||||||||||||||||:||||||||||||||||||||||||||||||||||||||||
Db       5534 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 5593

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5594 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 5653

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5654 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 5713

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5714 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 5773

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       5774 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 5833

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db       5834 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 5874


RESULT 2
AX377239
LOCUS       AX377239               13612 bp    DNA     linear   PAT 18-MAR-2002
DEFINITION  Sequence 1 from Patent WO0212562.
ACCESSION   AX377239
VERSION     AX377239.1  GI:19573528
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Kazemi,A., Kliem,S.E. and Koshy,B.
  TITLE     Haplotypes of the pla2g1b gene
  JOURNAL   Patent: WO 0212562-A 1 14-FEB-2002;
            Genaissance Pharmaceuticals, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .13612
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
     variation       3845
                     /note="PS1: polymorphic base G or A"
     variation       3968
                     /note="PS2: polymorphic base A or G"
     variation       6060
                     /note="PS2: polymorphic base A or G"
     variation       6844
                     /note="PS4: polymorphic base G or A"
     variation       9531
                     /note="PS5: polymorphic base G or A"
ORIGIN

  Query Match             99.9%;  Score 2497.4;  DB 6;  Length 13612;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2496;  Conservative    4;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6516 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 6575

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6576 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 6635

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6636 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 6695

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6696 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 6755

Qy        241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db       6756 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 6815

Qy        301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6816 ACCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 6875

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6876 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 6935

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6936 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 6995

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       6996 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 7055

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7056 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 7115

Qy        601 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7116 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 7175

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7176 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 7235

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7236 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 7295

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7296 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 7355

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7356 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 7415

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       7416 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 7475

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
```

```
Db      7476 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 7535
Qy      1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db      7536 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 7595
Qy      1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db      7596 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 7655
Qy      1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db      7656 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 7715
Qy      1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db      7716 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 7775
Qy      1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db      7776 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 7835
Qy      1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db      7836 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 7895
Qy      1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db      7896 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 7955
Qy      1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
Db      7956 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 8015
Qy      1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
Db      8016 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 8075
Qy      1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
Db      8076 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 8135
Qy      1621 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
Db      8136 TACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 8195
Qy      1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
Db      8196 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 8255
Qy      1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
Db      8256 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 8315
Qy      1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
Db      8316 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 8375
Qy      1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
Db      8376 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 8435
Qy      1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
Db      8436 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 8495
Qy      1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
Db      8496 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 8555
Qy      2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
Db      8556 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 8615
Qy      2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
Db      8616 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 8675
Qy      2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
Db      8676 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 8735
Qy      2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
Db      8736 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 8795
Qy      2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
Db      8796 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 8855
Qy      2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
Db      8856 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 8915
Qy      2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
Db      8916 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 8975
Qy      2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
Db      8976 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 9016
```

```
RESULT 3
AC003982/c
LOCUS       AC003982              122302 bp    DNA     linear   PRI 18-MAR-1999
DEFINITION  Homo sapiens PAC clone 166H1 from 12q, complete sequence.
ACCESSION   AC003982
VERSION     AC003982.1  GI:2769695
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 122302)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 122302)
  AUTHORS   Bradshaw,H., Wu,X. and Ozersky,P.
  TITLE     The sequence of Homo sapiens PAC clone 166H1
  JOURNAL   Unpublished (1999)
REFERENCE   3  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (13-JAN-1998) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
REFERENCE   4  (bases 1 to 122302)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (18-MAR-1999) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     SUBMITTED BY:  WUGSC
            Genome Sequencing Center
            Department of Genetics
            Washington University
            St. Louis MO 63108, USA
            http://genome.wustl.edu/gsc
            mailto:sapiens@watson.wustl.edu

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
```

between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:
This clone was originally isolated in the laboratory of Professor Graeme Bell, Howard Hughes Medical Institute and Departments of Biochemistry and Molecular Biology, and Medicine, The University of Chicago, Chicago, IL, USA. The clone was provided by the laboratory of Dr. Roger Cox at The Wellcome Trust Centre For Human Genetics, Oxford, UK. Some contig information was also obtained from Yamagata et al., Nature 384:455-8 (1996).

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-1, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is 278C19; the clone sequenced to the right is 15E1. Actual start of this clone is at base position 1 of 166H1; actual end is at 122302 of 166H1.

FEATURES Location/Qualifiers
source 1. .122302
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="12"
/map="12q"
/clone="166H1"
/clone_lib="RPCI-1"
repeat_region 1. .238
/rpt_family="Alu"
repeat_region 239. .265
/rpt_family="AT_rich"
repeat_region 268. .565
/rpt_family="Alu"
repeat_region 1001. .1309
/rpt_family="Alu"
repeat_region 1440. .1745
/rpt_family="Alu"
repeat_region 1909. .1940
/rpt_family="MER1_type"
repeat_region 1955. .1983
/rpt_family="AT_rich"
repeat_region 2063. .2362
/rpt_family="Alu"
repeat_region 2516. .2650
/rpt_family="Alu"
repeat_region 2674. .2730
/rpt_family="L1"
repeat_region 2735. .2788
/rpt_family="(GAAAA)n"
repeat_region 2790. .3077
/rpt_family="Alu"
repeat_region 3081. .3139
/rpt_family="L1"
repeat_region 3716. .4020
/rpt_family="Alu"
repeat_region 4196. .4492
/rpt_family="Alu"
gene 4560. .14034
/gene="WUGSC:H_166H1.1"
CDS join(4560. .5056,13418. .13748,13882. .14034)
/gene="WUGSC:H_166H1.1"
/note="unknown function; 60% similar to Z50177 (PID:g927403) (PID:g927402); H_166H1.1"
/codon_start=1
/evidence=not_experimental
/protein_id="AAB95634.1"
/db_xref="GI:2769696"
/translation="MKMSFALTFRSAKGRWIANPSQPCSKASIGLFVPASPPLDPEKV
KELQRFITLSKRLLVMTGAGISTESGIPDYRSEKVGLYARTDRRPIQHGDFVRSAPIR
QRYWARNFVGWPQFSSHQPNPAHWALSTWEKLGKLYWLVTQNVDALHTKAGSRRLTEL
HGCMDRAYCSVSVFLGSRVLCLDCGEQTPRGVLQERFQVLNPTWSAEAHGLAPDGDVF
LSEEQVRSFQVPTCVQCGGHLKPDVVFFGDTVNPDKVDFVHKRVKEADSLLVVGSSLQ
VYSGYRFILTAWEKKLPIAILNIGPTRSDDLACLKLNSRCGELLPLIDPC"
repeat_region 5366. .5655
/rpt_family="Alu"
repeat_region 5672. .5798
/rpt_family="Alu"
repeat_region 5802. .5834
/rpt_family="(TA)n"
repeat_region 5834. .5857
/rpt_family="POLY_A"
repeat_region 5870. .6184
/rpt_family="Alu"
repeat_region 6224. .6268
/rpt_family="(CA)n"
repeat_region 6269. .6558
/rpt_family="Alu"
repeat_region 6729. .7029
/rpt_family="Alu"
repeat_region 7051. .7172
/rpt_family="Alu"
repeat_region 7192. .7500
/rpt_family="Alu"
repeat_region 7518. .7820
/rpt_family="Alu"
repeat_region 8207. .8500
/rpt_family="Alu"
repeat_region 8576. .8878
/rpt_family="Alu"
repeat_region 9023. .9154
/rpt_family="Alu"
repeat_region 9165. .9470
/rpt_family="Alu"
repeat_region 9887. .10188
/rpt_family="Alu"
repeat_region 10221. .10342
/rpt_family="L1"
repeat_region 10408. .10707
/rpt_family="Alu"
repeat_region 10713. .11014
/rpt_family="Alu"
repeat_region 11020. .11110
/rpt_family="Alu"
repeat_region 11146. .11386
/rpt_family="Alu"
repeat_region 11208. .11386
/rpt_family="Alu"
repeat_region 11391. .11515
/rpt_family="Alu"
repeat_region 11568. .11705
/rpt_family="Alu"
repeat_region 11737. .12032
/rpt_family="Alu"
repeat_region 12034. .12066
/rpt_family="7SLRNA"
repeat_region 12075. .12372
/rpt_family="Alu"
repeat_region 12448. .12753
/rpt_family="Alu"
repeat_region 12763. .12874
/rpt_family="L2"

```
     repeat_region   12999. .13303
                     /rpt_family="Alu"
     misc_feature    complement(13674. .13752)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     misc_feature    complement(13840. .14195)
                     /note="match to EST N80020 (NID:g1242721) za91a08.s1"
     repeat_region   14196. .14229
                     /rpt_family="AT_rich"
     repeat_region   14240. .14339
                     /rpt_family="U6"
     repeat_region   14349. .14646
                     /rpt_family="Alu"
     repeat_region   14652. .14940
                     /rpt_family="Alu"

  Query Match             99.9%;  Score 2497.4;  DB 8;  Length 122302;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2495;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26288 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 26229

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26228 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 26169

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26168 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 26109

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      26108 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 26049

Qy        241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      26048 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 25989

Qy        301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||||:||||||||||||||||||||||||||||||
Db      25988 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 25929

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25928 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 25869

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25868 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 25809

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25808 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 25749

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25748 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 25689

Qy        601 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25688 GGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 25629

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25628 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 25569

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25568 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 25509

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db      25508 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 25449

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25448 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 25389

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25388 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 25329

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25328 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 25269

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:||||||||||||||||||
Db      25268 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 25209

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25208 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 25149

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25148 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 25089

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25088 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 25029

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      25028 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 24969

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24968 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 24909

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24908 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 24849

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24848 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 24789

Qy       1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24788 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 24729

Qy       1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24728 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 24669

Qy       1621 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24668 TACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGA 24609

Qy       1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24608 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 24549

Qy       1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24548 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 24489

Qy       1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24488 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 24429

Qy       1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24428 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 24369
```

```
Qy       1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24368 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 24309

Qy       1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24308 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 24249

Qy       2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24248 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 24189

Qy       2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 2160
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24188 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTTGGTTATAAAAGTAAACAAGC 24129

Qy       2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
             ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||||
Db      24128 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 24069

Qy       2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24068 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 24009

Qy       2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      24008 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 23949

Qy       2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23948 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 23889

Qy       2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      23888 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 23829

Qy       2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
             |||||||||||||||||||||||||||||||||||||||||
Db      23828 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 23788
```

```
RESULT 4
AC078926/c
LOCUS       AC078926               220384 bp    DNA     linear   HTG 26-MAR-2002
DEFINITION  Homo sapiens chromosome 12 clone RP11-836M11, WORKING DRAFT
            SEQUENCE, 11 unordered pieces.
ACCESSION   AC078926
VERSION     AC078926.23  GI:19718703
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 220384)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (11-AUG-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 220384)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Mar 26, 2002 this sequence version replaced gi:18449664.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
            -------------- Project Information
                Center project name: HCAM
                Center clone name: RP11-836M11
            -------------- Summary Statistics
                Sequencing vector: Plasmid;
                Sequencing vector: M13;
                Chemistry: Dye-primer Bodipy: 47% of reads
                Chemistry: Dye-terminator Big Dye: 53% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 234783 bases at least Q40
                Consensus quality: 241821 bases at least Q30
                Consensus quality: 245671 bases at least Q20
                Estimated insert size: 219187; sum-of-contigs estimation
                Quality coverage: 8.9x in Q20 bases; sum-of-contigs estimation
            --------------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 11 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     2156: contig of 2156 bp in length
            *     2157     2256: gap of unknown length
            *     2257     7112: contig of 4856 bp in length
            *     7113     7212: gap of unknown length
            *     7213    13283: contig of 6071 bp in length
```

```
            *    13284    13383: gap of unknown length
            *    13384    20660: contig of 7277 bp in length
            *    20661    20760: gap of unknown length
            *    20761    34473: contig of 13713 bp in length
            *    34474    34573: gap of unknown length
            *    34574    44394: contig of 9821 bp in length
            *    44395    44494: gap of unknown length
            *    44495    57126: contig of 12632 bp in length
            *    57127    57226: gap of unknown length
            *    57227    80266: contig of 23040 bp in length
            *    80267    80366: gap of unknown length
            *    80367   109398: contig of 29032 bp in length
            *   109399   109498: gap of unknown length
            *   109499   149812: contig of 40314 bp in length
            *   149813   149912: gap of unknown length
            *   149913   220384: contig of 70472 bp in length.
FEATURES             Location/Qualifiers
     source          1. .220384
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12"
                     /clone="RP11-836M11"
     gap             2157. .2256
                     /estimated_length=unknown
     gap             7113. .7212
                     /estimated_length=unknown
     gap             13284. .13383
                     /estimated_length=unknown
     gap             20661. .20760
                     /estimated_length=unknown
     gap             34474. .34573
                     /estimated_length=unknown
     gap             44395. .44494
                     /estimated_length=unknown
     gap             57127. .57226
                     /estimated_length=unknown
     gap             80267. .80366
                     /estimated_length=unknown
     gap             109399. .109498
                     /estimated_length=unknown
     gap             149813. .149912
                     /estimated_length=unknown
ORIGIN

  Query Match           99.9%;  Score 2497.4;  DB 14;  Length 220384;
  Best Local Similarity  99.8%;  Pred. No. 0;
  Matches 2495;  Conservative    5;  Mismatches    1;  Indels    0;  Gaps    0;

Qy          1 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71511 GTCGTGTCACTGCTGTCCAGCCTGGGTAACAGAGCAACTCTGTCTCAAAAAAAAAAAATG 71452

Qy         61 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71451 CTTTCAATAAATATATGATAAAAGGACTTATATTTTTTCAAGCCATAGGATCATTTCTCC 71392

Qy        121 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71391 TGAAGCATCTTGGCGAAGTCATCCCCACCTGTTCCTGAGAGTGGGCAGGTGAGGGCTGAC 71332

Qy        181 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71331 CTATTGCTCTGCACTTACTCCTATCTCAGCTGTCCCTCCCACTTTCCAGGTGCTGCCAGA 71272

Qy        241 CACATGACAACTGCTACGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 300
              |||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||
Db      71271 CACATGACAACTGCTATGACCAGGCCAAGAAGCTGGACAGCTGTAAATTTCTGCTGGACA 71212

Qy        301 MMCCGTACACCCACACCTATTCATACTCRTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 360
              ::||||||||||||||||||||||||:||||||||||||||||||||||||||||||||
Db      71211 ACCCGTACACCCACACCTATTCATACTCGTGCTCTGGCTCGGCAATCACCTGTAGCAGTA 71152

Qy        361 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71151 GGTTTATCCCTTCCTTGACCTATGAATTCTAGTTGGTTCTCAGTAGGCCGGGGGGAAATA 71092

Qy        421 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71091 ATAGTAACAACAGCCATGATTTAGTGTTAATTTTCTTGGTTCTGGGCAGTGTCTCCTTTA 71032

Qy        481 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      71031 ATCCTCAGAACAACACTATGGGATAGGTACAATTATCCTCACTTAACAGATAAGAAAACT 70972

Qy        541 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70971 GAGGCTCAGAAGGCTGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTT 70912

Qy        601 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70911 GGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCAC 70852

Qy        661 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70851 AGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACC 70792

Qy        721 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70791 TCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAA 70732

Qy        781 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70731 ATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTT 70672

Qy        841 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70671 CAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATG 70612

Qy        901 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70611 TGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAG 70552

Qy        961 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70551 GGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAA 70492

Qy       1021 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db      70491 CCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGCGATTCTCCTGCCTAAGC 70432

Qy       1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70431 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 70372

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70371 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 70312

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70311 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 70252

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70251 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 70192

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70191 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 70132

Qy       1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      70131 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 70072

Qy       1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 1500
```

```
                 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       70071 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTC 70012

Qy        1501 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 1560
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       70011 ATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTC 69952

Qy        1561 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 1620
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69951 CACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAA 69892

Qy        1621 TACACACACACACACACACTCACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAGA 1680
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69891 TACACACACACACACACACTCACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAGA 69832

Qy        1681 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 1740
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69831 AACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGA 69772

Qy        1741 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 1800
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69771 CTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGA 69712

Qy        1801 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 1860
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69711 AGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATT 69652

Qy        1861 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 1920
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69651 AGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAA 69592

Qy        1921 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 1980
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69591 TTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGC 69532

Qy        1981 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 2040
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69531 CTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAG 69472

Qy        2041 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 2100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69471 AGAAACAAAACTAATAAGATTCCTGAAGGTAAGCAGAGATACGTAAATTATATGTAATAA 69412

Qy        2101 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTGGTTATAAAAGTAAACAAGC 2160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69411 AGTTTAAATGCATTTTAACTGTAATCTTATTGTTTATTTGGTTATAAAAGTAAACAAGC 69352

Qy        2161 CAAAAGTAATGCAACTTCAAACKCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 2220
               ||||||||||||||||||||||:|||||||||||||||||||||||||||||||||||||
Db       69351 CAAAAGTAATGCAACTTCAAACTCTACATAAATATCTATTATGGAAAGTGGAAGGCATCT 69292

Qy        2221 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 2280
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69291 ATAATCCTACTACCCAAAGATAACCAGTTACATATTCCTCCAGATTTTTGGGGCATACAC 69232

Qy        2281 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 2340
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69231 TAGCTTTTTTTATTTGGGAAAATTTCCATGTGCAGGCATACCTAATTTTTCTAAATGTCT 69172

Qy        2341 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 2400
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69171 ATGTAGTATTCCATTTAAGGATGTTCCATAATTTTTAAAATACATGCTTTAAAGTAGAGA 69112

Qy        2401 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 2460
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       69111 AACTAGGTTGGGCATGGTGGCTCACGCCTGTATCCCAGCACTTTGGGAGGCCGAGGCAAA 69052

Qy        2461 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 2501
               |||||||||||||||||||||||||||||||||||||||||
Db       69051 TGGATCACTTGAGGTCCGGAGTTTGAGACCAGCCTGGACAA 69011
```

RESULT 5

```
AC073930
LOCUS       AC073930               189729 bp    DNA     linear   HTG 01-MAR-2002
DEFINITION  Homo sapiens chromosome 12q clone RP11-144B2, WORKING DRAFT
            SEQUENCE, 8 unordered pieces.
ACCESSION   AC073930
VERSION     AC073930.23  GI:18958593
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 189729)
  AUTHORS   Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,
            Alsbrooks,S.L., Amaratunge,H.C., Are,J.R., Ayele,M., Banks,T.,
            Barbaria,J., Benton,J., Bimage,K., Blankenburg,K., Bonnin,D.,
            Bouck,J., Bowie,S., Brieva,M., Brown,E., Brown,M., Bryant,N.P.,
            Buhay,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,
            Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,
            Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,
            Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,
            Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,
            Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H.,
            Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,
            Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,
            Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,
            Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,
            Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,
            Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,
            Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,
            Homsi,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,
            Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,
            Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,
            Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,
            Li,J., Li,Z., Lichtarge,O., Lieu,C., Liu,J., Liu,W., Loulseged,H.,
            Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,
            Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,
            Massey,E., Mawhiney,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,
            Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,
            Moser,M., Neal,D., Newtson,J., Newtson,N., Nguyen,A., Nguyen,N.,
            Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okwuonu,G.,
            Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,
            Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,Y.,
            Rives,M., Rojas,A., Rojubokan,I., Rolfe,M., Ruiz,S., Savery,G.,
            Scherer,S., Scott,G., Shen,H., Shooshtari,N., Sisson,I.,
            Sodergren,E., Sonaike,T., Sparks,A., Stanley,H., Stone,H.,
            Sutton,A., Svatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,
            Tansey,J., Taylor,C., Taylor,T., Telfrod,B., Thomas,N., Thomas,S.,
            Usmani,K., Vasquez,L., Vera,V., Villalon,D., Vinson,R., Wang,Q.,
            Wang,S., Ward-Moore,S., Warren,R., Washington,C., Watlington,S.,
            Williams,G., Williamson,A., Wleczyk,R., Wooden,S., Worley,K.,
            Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,
            Weinstock,G. and Gibbs,R.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (05-JUL-2000) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
REFERENCE   3  (bases 1 to 189729)
  AUTHORS   Worley,K.C.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) Human Genome Sequencing Center, Department
            of Molecular and Human Genetics, Baylor College of Medicine, One
            Baylor Plaza, Houston, TX 77030, USA
COMMENT     On Feb 27, 2002 this sequence version replaced gi:14861669.
            -------------- Genome Center
                Center: Baylor College of Medicine
                Center code: BCM
                Web site: http://www.hgsc.bcm.tmc.edu/
                Contact: hgsc-help@bcm.tmc.edu
```

```
            ---------- Project Information
                Center project name: HBWS
                Center clone name: RP11-144B2
            ---------- Summary Statistics
                Sequencing vector: Plasmid; M77789
                Sequencing vector: M13; L08821
                Chemistry: Dye-terminator Big Dye: 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 191640 bases at least Q40
                Consensus quality: 195135 bases at least Q30
                Consensus quality: 197786 bases at least Q20
                Estimated insert size: 193340; sum-of-contigs estimation
                Quality coverage: 0x in Q20 bases; agarose-fp estimation
                Quality coverage: 11.7x in Q20 bases; sum-of-contigs estimation
            ----------
            * NOTE: Estimated insert size may differ from sequence length
            *   (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 8 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *       1    64773: contig of 64773 bp in length
            *   64774    64873: gap of unknown length
            *   64874    99097: contig of 34224 bp in length
            *   99098    99197: gap of unknown length
            *   99198   135331: contig of 36134 bp in length
            *  135332   135431: gap of unknown length
            *  135432   161344: contig of 25913 bp in length
            *  161345   161444: gap of unknown length
            *  161445   171231: contig of 9787 bp in length
            *  171232   171331: gap of unknown length
            *  171332   180922: contig of 9591 bp in length
            *  180923   181022: gap of unknown length
            *  181023   187271: contig of 6249 bp in length
            *  187272   187371: gap of unknown length
            *  187372   189729: contig of 2358 bp in length.
FEATURES             Location/Qualifiers
     source          1. .189729
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="12q"
                     /clone="RP11-144B2"
     gap             64774. .64873
                     /estimated_length=unknown
     gap             99098. .99197
                     /estimated_length=unknown
     gap             135332. .135431
                     /estimated_length=unknown
     gap             161345. .161444
                     /estimated_length=unknown
     gap             171232. .171331
                     /estimated_length=unknown
     gap             180923. .181022
                     /estimated_length=unknown
     gap             187272. .187371
                     /estimated_length=unknown
ORIGIN

  Query Match             84.5%;  Score 2111.2;  DB 14;  Length 189729;
  Best Local Similarity   99.9%;  Pred. No. 0;
  Matches 2110;  Conservative    2;  Mismatches    0;  Indels    0;  Gaps    0;

Qy        390 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 449
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171332 TAGTTGGTTCTCAGTAGGCCGGGGGGAAATAATAGTAACAACAGCCATGATTTAGTGTTA 171391

Qy        450 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 509
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Db     171392 ATTTTCTTGGTTCTGGGCAGTGTCTCCTTTAATCCTCAGAACAACACTATGGGATAGGTA 171451

Qy        510 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 569
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171452 CAATTATCCTCACTTAACAGATAAGAAAACTGAGGCTCAGAAGGCTGAGCTATTTGCCCA 171511

Qy        570 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 629
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171512 AGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACA 171571

Qy        630 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 689
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171572 GGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCT 171631

Qy        690 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 749
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171632 CAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAG 171691

Qy        750 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 809
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171692 CGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTAC 171751

Qy        810 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 869
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171752 GTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAA 171811

Qy        870 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 929
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171812 AGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATC 171871

Qy        930 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 989
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171872 TTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTT 171931

Qy        990 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 1049
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     171932 TTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCC 171991

Qy       1050 TCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 1109
              ||||||||||||:||||||||||||||||||||||||||||||||||||||||||||||
Db     171992 TCCCGGGTTCAAGCGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGC 172051

Qy       1110 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 1169
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172052 ACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCC 172111

Qy       1170 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 1229
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172112 CAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCT 172171

Qy       1230 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 1289
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172172 GGAATTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTAT 172231

Qy       1290 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 1349
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172232 ATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAAT 172291

Qy       1350 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 1409
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172292 GCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAG 172351

Qy       1410 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 1469
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172352 AATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATAT 172411

Qy       1470 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 1529
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172412 TGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATA 172471

Qy       1530 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 1589
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172472 TGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTG 172531
```

```
Qy       1590 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 1649
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172532 ATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAG 172591

Qy       1650 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 1709
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172592 TTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATAT 172651

Qy       1710 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 1769
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172652 TAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAA 172711

Qy       1770 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 1829
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172712 GTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTG 172771

Qy       1830 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 1889
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172772 AAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGT 172831

Qy       1890 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 1949
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172832 CCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCA 172891

Qy       1950 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 2009
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172892 GTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAA 172951

Qy       2010 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 2069
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     172952 AAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTCCTGAAGG 173011

Qy       2070 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 2129
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173012 TAAGCAGAGATACGTAAATTATATGTAATAAAGTTTAAATGCATTTTAACTGTAATCTTA 173071

Qy       2130 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACKCTACAT 2189
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||:|||||||
Db     173072 TTGTTTATTTTGGTTATAAAAGTAAACAAGCCAAAAGTAATGCAACTTCAAACGCTACAT 173131

Qy       2190 AAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTT 2249
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173132 AAATATCTATTATGGAAAGTGGAAGGCATCTATAATCCTACTACCCAAAGATAACCAGTT 173191

Qy       2250 ACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCAT 2309
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173192 ACATATTCCTCCAGATTTTTGGGGCATACACTAGCTTTTTTTATTTGGGAAAATTTCCAT 173251

Qy       2310 GTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCAT 2369
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173252 GTGCAGGCATACCTAATTTTTCTAAATGTCTATGTAGTATTCCATTTAAGGATGTTCCAT 173311

Qy       2370 AATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCT 2429
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173312 AATTTTTAAAATACATGCTTTAAAGTAGAGAAACTAGGTTGGGCATGGTGGCTCACGCCT 173371

Qy       2430 GTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGAC 2489
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     173372 GTATCCCAGCACTTTGGGAGGCCGAGGCAAATGGATCACTTGAGGTCCGGAGTTTGAGAC 173431

Qy       2490 CAGCCTGGACAA 2501
             ||||||||||||
Db     173432 CAGCCTGGACAA 173443


RESULT 6
AC068981/c
LOCUS       AC068981              178887 bp    DNA     linear   HTG 01-SEP-2000
DEFINITION  Homo sapiens chromosome 15 clone RP11-580I7, WORKING DRAFT
            SEQUENCE, 37 unordered pieces.
ACCESSION   AC068981
VERSION     AC068981.3  GI:9958291
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     The sequence of Homo sapiens clone
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 178887)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (15-MAY-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
COMMENT     On Sep 1, 2000 this sequence version replaced gi:8748938.

            ---------------- Genome Center -----------------
            Center: Washington University Genome Sequencing Center
            Center code: WUGSC
            Web site:http://genome.wustl.edu/gsc/index.shtml
            -------------- Project Information --------------
            Center project name: H_NH0580I07
            -------------- Summary Statistics  --------------
            Sequencing vector: M13; 100%
            Sequencing vector: plasmid; 0%
            Chemistry: Dye-primer ET; 100% of reads
            Chemistry: Dye-terminator Big Dye; 0% of reads
            Assembly program: Phrap; version 0.990319
            Consensus quality: 157045 bases at least Q40
            Consensus quality: 163476 bases at least Q30
            Consensus quality: 167183 bases at least Q20
            Insert size: 192000; agarose-fp
            Insert size: 174828; sum-of-contigs
            Quality coverage: 3.01 in Q20 bases; agarose-fp
            Quality coverage: 3.40 in Q20 bases; sum-of-contigs
            ------------------------------------------------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 37 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *        1     1242: contig of 1242 bp in length
            *     1243     1342: gap of unknown length
            *     1343     2435: contig of 1093 bp in length
            *     2436     2535: gap of unknown length
            *     2536     3664: contig of 1129 bp in length
            *     3665     3764: gap of unknown length
            *     3765     4864: contig of 1100 bp in length
            *     4865     4964: gap of unknown length
            *     4965     7695: contig of 2731 bp in length
            *     7696     7795: gap of unknown length
            *     7796    10187: contig of 2392 bp in length
            *    10188    10287: gap of unknown length
            *    10288    12303: contig of 2016 bp in length
            *    12304    12403: gap of unknown length
            *    12404    14390: contig of 1987 bp in length
            *    14391    14490: gap of unknown length
            *    14491    16285: contig of 1795 bp in length
            *    16286    16385: gap of unknown length
            *    16386    18364: contig of 1979 bp in length
            *    18365    18464: gap of unknown length
            *    18465    20794: contig of 2330 bp in length
            *    20795    20894: gap of unknown length
            *    20895    23473: contig of 2579 bp in length
            *    23474    23573: gap of unknown length
            *    23574    25387: contig of 1814 bp in length
            *    25388    25487: gap of unknown length
            *    25488    28870: contig of 3383 bp in length
```

```
            *    28871    28970: gap of unknown length
            *    28971    32203: contig of 3233 bp in length
            *    32204    32303: gap of unknown length
            *    32304    32762: contig of 459 bp in length
            *    32763    32862: gap of unknown length
            *    32863    36428: contig of 3566 bp in length
            *    36429    36528: gap of unknown length
            *    36529    39558: contig of 3030 bp in length
            *    39559    39658: gap of unknown length
            *    39659    44249: contig of 4591 bp in length
            *    44250    44349: gap of unknown length
            *    44350    48970: contig of 4621 bp in length
            *    48971    49070: gap of unknown length
            *    49071    54728: contig of 5658 bp in length
            *    54729    54828: gap of unknown length
            *    54829    60549: contig of 5721 bp in length
            *    60550    60649: gap of unknown length
            *    60650    66747: contig of 6098 bp in length
            *    66748    66847: gap of unknown length
            *    66848    73820: contig of 6973 bp in length
            *    73821    73920: gap of unknown length
            *    73921    80722: contig of 6802 bp in length
            *    80723    80822: gap of unknown length
            *    80823    86541: contig of 5719 bp in length
            *    86542    86641: gap of unknown length
            *    86642    91906: contig of 5265 bp in length
            *    91907    92006: gap of unknown length
            *    92007    99976: contig of 7970 bp in length
            *    99977   100076: gap of unknown length
            *   100077   108154: contig of 8078 bp in length
            *   108155   108254: gap of unknown length
            *   108255   114943: contig of 6689 bp in length
            *   114944   115043: gap of unknown length
            *   115044   122680: contig of 7637 bp in length
            *   122681   122780: gap of unknown length
            *   122781   130297: contig of 7517 bp in length
            *   130298   130397: gap of unknown length
            *   130398   138520: contig of 8123 bp in length
            *   138521   138620: gap of unknown length
            *   138621   146448: contig of 7828 bp in length
            *   146449   146548: gap of unknown length
            *   146549   152895: contig of 6347 bp in length
            *   152896   152995: gap of unknown length
            *   152996   162254: contig of 9259 bp in length
            *   162255   162354: gap of unknown length
            *   162355   178887: contig of 16533 bp in length.
FEATURES             Location/Qualifiers
     source          1. .178887
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /clone="RP11-580I7"
     misc_feature    1. .1242
                     /note="assembly_name:Contig22"
     gap             1243. .1342
                     /estimated_length=unknown
     misc_feature    1343. .2435
                     /note="assembly_name:Contig24"
     gap             2436. .2535
                     /estimated_length=unknown
     misc_feature    2536. .3664
                     /note="assembly_name:Contig25"
     gap             3665. .3764
                     /estimated_length=unknown
     misc_feature    3765. .4864
                     /note="assembly_name:Contig27"
     gap             4865. .4964
                     /estimated_length=unknown
     misc_feature    4965. .7695
                     /note="assembly_name:Contig28"
     gap             7696. .7795
                     /estimated_length=unknown
     misc_feature    7796. .10187
                     /note="assembly_name:Contig30"
     gap             10188. .10287
                     /estimated_length=unknown
     misc_feature    10288. .12303
                     /note="assembly_name:Contig31"
     gap             12304. .12403
                     /estimated_length=unknown
     misc_feature    12404. .14390
                     /note="assembly_name:Contig32"
     gap             14391. .14490
                     /estimated_length=unknown
     misc_feature    14491. .16285
                     /note="assembly_name:Contig33"
     gap             16286. .16385
                     /estimated_length=unknown
     misc_feature    16386. .18364
                     /note="assembly_name:Contig34"
     gap             18365. .18464
                     /estimated_length=unknown
     misc_feature    18465. .20794
                     /note="assembly_name:Contig35"
     gap             20795. .20894
                     /estimated_length=unknown
     misc_feature    20895. .23473
                     /note="assembly_name:Contig36"
     gap             23474. .23573
                     /estimated_length=unknown
     misc_feature    23574. .25387
                     /note="assembly_name:Contig37"
     gap             25388. .25487
                     /estimated_length=unknown
     misc_feature    25488. .28870
                     /note="assembly_name:Contig38"
     gap             28871. .28970
                     /estimated_length=unknown
     misc_feature    28971. .32203
                     /note="assembly_name:Contig39
                     clone_end:SP6
                     vector_side:right"
     gap             32204. .32303
                     /estimated_length=unknown
     misc_feature    32304. .32762
                     /note="assembly_name:Contig12"
     gap             32763. .32862
                     /estimated_length=unknown
     misc_feature    32863. .36428
                     /note="assembly_name:Contig40"
     gap             36429. .36528

  Query Match             18.5%;  Score 462.6;  DB 14;  Length 178887;
  Best Local Similarity   62.1%;  Pred. No. 1.3e-83;
  Matches  890;  Conservative    1;  Mismatches  485;  Indels   57;  Gaps    8;

Qy        600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
              || || |||| ||||||||||||      | | ||  |||||||| |||||||  |||
Db      90620 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 90561

Qy        660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
              ||||||  | | |   | ||||||||  |||  ||||||| |||||||  |||  || ||
Db      90560 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 90501

Qy        720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
              |||||| ||||||||||||||||||||  | |  || ||||| || | |||||||
Db      90500 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 90441

Qy        780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
              |  |||| |||||||||| |||| || | | |||| ||| |||  | | |||||||||
Db      90440 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 90383

Qy        840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
              |||| | |||| |||  ||  || |||||||||||| |||||||||||| ||||||||||
Db      90382 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 90324
```

```
Qy        900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
Db      90323 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 90264

Qy        960 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
Db      90263 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTGTGGGACTGCCACTGGAGTGCAGTG 90205

Qy       1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
Db      90204 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 90145

Qy       1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
Db      90144 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 90085

Qy       1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
Db      90084 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 90025

Qy       1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
Db      90024 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 89965

Qy       1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
Db      89964 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 89913

Qy       1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
Db      89912 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 89867

Qy       1380 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
Db      89866 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 89828

Qy       1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
Db      89827 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 89768

Qy       1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
Db      89767 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 89708

Qy       1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
Db      89707 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 89653

Qy       1620 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
Db      89652 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 89593

Qy       1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
Db      89592 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 89533

Qy       1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
Db      89532 TTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 89475

Qy       1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
Db      89474 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 89415

Qy       1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
Db      89414 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 89355

Qy       1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
Db      89354 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACGCCACTATACTC 89295

Qy       1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 2029
Db      89294 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAACA 89242
```

```
RESULT 7
AC009244
LOCUS       AC009244               190162 bp    DNA     linear   PRI 26-MAY-2002
DEFINITION  Homo sapiens chromosome 7 clone RP11-305M3, complete sequence.
ACCESSION   AC009244
VERSION     AC009244.25  GI:21217399
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 190162)
  AUTHORS   Bubb,K.L., Desmarais,C.L., Ramsey,S.A. and Hubley,R.M.
  TITLE     Direct Submission
  JOURNAL   Submitted (07-AUG-1999) Human Genome Center, University of
            Washington, Box 352145, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Zhou,Y., James,R.A., Raymond,C., Haugen,E.D. and
            Olson,M.V.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-OCT-2000) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
REFERENCE   4  (bases 1 to 190162)
  AUTHORS   Kaul,R.K., Olson,M.V., Zhou,Y., James,R.A., Rouse,G., Wu,Z.,
            Saenphimmachak,C., Phelps,K.A., Buckley,D., Kibukawa,M., Raymond,C.
            and Haugen,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (26-MAY-2002) Genome Center, University of Washington,
            Box 352145, Seattle, WA 98195, USA
COMMENT     On May 26, 2002 this sequence version replaced gi:11038533.
            -------------- Genome Center
                Center: University of Washington Genome Center
                Center Code: UWGC
                Web site: http://www.genome.washington.edu
                Contact: uwgchtgs@u.washington.edu
            -------------- Project Information
                Center project name: chr-7
                Center clone name: RP11-305M3 (djs195)
            ------------- Summary Statistics
                Assembly program: Phrap; version 0.990319
                Consensus quality: 189969 bases at least Q40
                Consensus quality: 190110 bases at least Q30
                Consensus quality: 190161 bases at least Q20
                Insert size: 190162; sum-of-contigs
                Quality coverage: 9.7x in Q20 bases; sum-of-contigs
            --------------.
            Overlapping Sequences:
                5':  RP11-382M23 (UWGC:djs734g) AC093149 1340-bp overlap
                3':  RP11-448A19 (UWGC:djs705) AC078846 8765-bp overlap
            --------------------
            Sequence Quality Assessment:
             This entry has been annotated with sequence quality
             estimates computed by the Phrap assembly program.
             All manually edited bases have been reduced to quality zero.
             Quality levels above 40 are expected to have less than
             1 error in 10,000 bp.
             Base-by-base quality values are not generally visible from the
             GenBank flat file format but are available as part
             of this entry's ASN.1 file.

            This sequence was finished as follows unless otherwise noted:
```

all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., Phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest.

Sequence Validation:

This sequence has been validated by Multiple Complete Digest fingerprinting. Comparison of the experimentally derived digest fragments with sequence-predicted fragments is given below. The electronically-digested sequence consists of both insert and vector, in order to accurately represent the entire circular BAC. Small fragments below a variable cutoff (approximately 400-800 bp) are not resolved in the fingerprint and hence do not appear in the table. There are no significant remaining discrepancies between the experimental and predicted values. Uniquely ordered fragments are separated by dashed lines.

NsiI		BglII		EcoRI	
SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt	SeqDerMap	FngrPrnt
21251	20929	12417	12174	8696	8794
12148	11918	2067	2041	6	<800
199	<800	5666	5826	1775	1773
3275	3221	254	<800	2000	1991
20379	20929	3818	3836	6313	6324
15555	15150	747	735	3256	3226
675	<800	9196	9237	325	<800
5493	5475	4154	4101	12989	12723
516	<800	5820	5826	5955	5951
5810	5831	118	<800	611	<800
9488	9435	2524	2359	1995	1991
2908	2935	8844	8768	11945	12009
779	840	332	<800	5577	5593
821	840	564	<800	4139	4124
315	<800	26	<800	955	940
3640	3671	4675	4644	2961	2940
428	<800	1801	1803	5264	5205
1144	1119	117	<800	1535	1534
4684	4717	6797	6799	3313	3383
20	<800	3146	3177	3760	3761
492	<800	2376	2359	3061	3110
10935	10879	210	<800	9605	9505
5921	6010	5135	5110	335	<800
529	<800	744	735	2190	2194
4354	4384	881	947	5117	5205
1311	1293	413	<800	13558	13434
10	<800	6389	6456	9367	9505
374	<800	296	<800	8080	8019
1497	1480	7160	7119	4276	4243
4079	4083	4846	4822	7830	8019
337	<800	2914	2924	231	<800
4483	4384	3803	3836	533	<800
1530	1480	1832	1803	668	<800
249	<800	674	<800	12314	12009
8831	8995	2515	2513	18561	18820
244	<800	2163	2137	11006	11013
1019	1000	466	<800	8756	8794
841	840	927	947	7	<800
8114	8086	5963	5826		
9184	9435	7440	7660		
7121	7263	5159	4822		
1122	1119	677	<800		
49	<800	1399	1373		
11928	11918	1565	1554		
1499	1480	2388	2513		
2585	2575	4826	4822		
699	<800	7661	7660		
		3624	3705		
		8693	8768		
		1749	1803		
		1642	1635		
		5295	5362		
		10628	10715		
		3227	3177		
		589	<800		

Query Match 18.5%; Score 462.6; DB 8; Length 190162;
Best Local Similarity 61.3%; Pred. No. 1.3e-83;
Matches 900; Conservative 1; Mismatches 520; Indels 46; Gaps 8;

```
Qy      604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
            |||| ||| ||     |||| || | |||||||| |||||||||| | | ||||
Db    31687 TTTTCCTTTTTTTAAAAGTAGAAACGGAGTCTTGCTGGGTCACCCAGGCTGGGGTGCAGT 31746

Qy      664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
            ||||| | |  |||||||| ||| | |||| | || | |||| | || |||
Db    31747 AGTGCAGTCGTGTTTCACTGCATCCTTCAATTCCTAGGAACAGGTGATCCTCCCACTTCA 31806
```

```
Qy       724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db     31807 GTCTTCTGAGTAGCTGGGACTTCAGGCATGTGCCACCATGCCCAGCTAATTTAAAAAAAA 31866

Qy       784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db     31867 TGTTTGAAGACACAGGGTCTTGCTGTGTTGCTCAGGCTGGTCTCAAACAGCCAGCCTCAA 31926

Qy       844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db     31927 GTGATCCTCCTGCCTCATCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCATGCCC 31986

Qy       904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
Db     31987 AGCCCATTTCTTTTTTCTGTTTTTTTTTTAAATTTAATTTAATTTTTATTTTTTATTTT 32046

Qy       964 TTACTGCCTATTTTCTTCCTT-TCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db     32047 TATTTTTATTTTTATTTATTTATTTTTTTTGAGATGAAGTCTCACTCTGTCGCCCAGGCT 32106

Qy      1023 TCCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTC 1068
Db     32107 GGAGTGCAGTGGCATGATCTCGGCTCACTGAAACCTCTGCCTCCCAGGTTCAAGCGATTC 32166

Qy      1069 TCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAAT 1128
Db     32167 TCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCCCTCCACTATACCCGGCTAA- 32225

Qy      1129 TTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCC 1188
Db     32226 TTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGTCAGGCTGGTCTCGAACTCC 32285

Qy      1189 TGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTC 1248
Db     32286 TGACCTCAGATGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCC 32345

Qy      1249 ACTGTGCCTGGCCGATTACTGTCTATTTT---CTTTATTGCTATATCCCCAGATCTAGAG 1305
Db     32346 ACCGCACCCGGCTGCCCCATTTTTTTTTTAAACAAATGCTTACTGACCTACGCTAAGAG 32405

Qy      1306 CAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATA 1365
Db     32406 CTGTAAGGGCAAATGAGACAATGTCCTTACCCTCAGAATCTTATATTTTGGTGAGATAAT 32465

Qy      1366 AACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTG 1425
Db     32466 ACACATATATAAATAATGATGATACAAGGACATATATAATAAATGCCATGACAGATTATA 32525

Qy      1426 CATTCTGCTTTTTCAC-TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTA 1484
Db     32526 AATCCATATTCACTACACAAACAAGAGGCTTAGGAAAGTTCAGATTCCCACTTTGGGAAG 32585

Qy      1485 TTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTT 1544
Db     32586 CTGAGGCAGGTGGATCACCTGAGGTCAGGAATTCA--AGACCAGCGTGGCCAACATGGTG 32643

Qy      1545 ATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAAC 1604
Db     32644 AAACCCTGTCTTTACTAAGAGTACAAAAATTAGCTGAGTGTGGTGGTGCAGGCCTGTAAT 32703

Qy      1605 CCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTT 1664
Db     32704 CCTAGCTACTCATGAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCGGAGGCTGCA 32763

Qy      1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTG---------------------- 1702
Db     32764 GTGAGCCGAGATCGCGCCACTGCACTCTAGCCTGGGTGACAGAGTGAGACTCCATCTCAA 32823

Qy      1703 TTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTAC 1762
Db     32824 AAAAAAAAAAAAAAAAAAAAAGAGGCCAGGCGCGGTGGCTCACACCTGTAATCCCAGCAC 32883

Qy      1763 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAA 1822
Db     32884 TTTGGGAGCCTGAGGTGGGCGGATCA--TGAGGTCAGAAGTTCGAGACCAGCCTGGCCAA 32941

Qy      1823 TATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGC 1882
Db     32942 TATGGTGAAACCCCGTCTTTACTAAAAGTACAAAAATTAGCTGGGCGTGGTGCCACGCAC 33001

Qy      1883 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 1942
Db     33002 CTGTAGTCCCAGCTACTCAGGAGGCTAAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGA 33061

Qy      1943 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCT 2002
Db     33062 GGTTGCAGTGAGCCGAGATTATGCCACTGCACTCCAGCCTGGGTGACAGAGCAAGACTCC 33121

Qy      2003 ATCTCAAAAAAAATAAATAAATAAAATA 2029
Db     33122 GTCTCAAAAAAAAAAAAGAAAAGAAAA 33148
```

RESULT 8
AP000589

```
LOCUS       AP000589              123192 bp    DNA     linear   HTG 30-MAY-2000
DEFINITION  Homo sapiens chromosome 11 clone CMB9-105N5 map 11q13, WORKING
            DRAFT SEQUENCE, 19 unordered pieces.
ACCESSION   AP000589
VERSION     AP000589.3  GI:8118795
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens 123,192 genomic DNA of 11q13
  JOURNAL   Published Only in DataBase (1999)
REFERENCE   2  (bases 1 to 123192)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-OCT-1999) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            Kitasato Univ., 1-15-1 Kitasato, Sagamihara, Kanagawa 228-8555,
            Japan (E-mail:hattori@gsc.riken.go.jp,
            URL:http://hgp.gsc.riken.go.jp/, Tel:81-42-778-9923,
            Fax:81-42-778-9924)
COMMENT     On May 31, 2000 this sequence version replaced gi:6997479.
            -------------- Genome Center
                Center: RIKEN Genomic Sciences Center(GSC)
                Center code: RIKEN
                Web site: http://hgp.gsc.riken.go.jp/
                Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
                Center project name: HumDraft11
                Center clone name: CMB9-105N5
            -------------- Summary Statistics
                Sequencing vector: PCR products; 100% of reads
                Chemistry: Dye-terminator ET-amersham; 100% of reads
                Assembly program: Phrap; version 0.990329
                Consensus quality: 115116 bases at least Q40
                Consensus quality: 118340 bases at least Q30
                Consensus quality: 120191 bases at least Q20
                Insert size: 121392; sum-of-contigs
                Quality coverage: 5.08x in Q20 bases; sum-of-contigs
            --------------.
            NOTE: This is a 'working draft' sequence. It currently consists of
            19 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
```

```
as soon as it is available and the accession number will be
preserved
         1     15827 contig of  15827 bp in length
     15928     30484 contig of  14557 bp in length
     30585     40602 contig of  10018 bp in length
     40703     51912 contig of  11210 bp in length
     52013     60022 contig of   8010 bp in length
     60123     70438 contig of  10316 bp in length
     70539     77945 contig of   7407 bp in length
     78046     85717 contig of   7672 bp in length
     85818     92399 contig of   6582 bp in length
     92500     98215 contig of   5716 bp in length
     98316    104148 contig of   5833 bp in length
    104249    109166 contig of   4918 bp in length
    109267    112457 contig of   3191 bp in length
    112558    115076 contig of   2519 bp in length
    115177    117063 contig of   1887 bp in length
    117164    118652 contig of   1489 bp in length
    118753    120425 contig of   1673 bp in length
    120526    121843 contig of   1318 bp in length
    121944    123192 contig of   1249 bp in length
Sequence updated (18-Oct-1999)
Sequence updated (26-May-2000).
* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*         1     15827: contig of 15827 bp in length
*     15828     15927: gap of 100 bp
*     15928     30484: contig of 14557 bp in length
*     30485     30584: gap of 100 bp
*     30585     40602: contig of 10018 bp in length
*     40603     40702: gap of 100 bp
*     40703     51912: contig of 11210 bp in length
*     51913     52012: gap of 100 bp
*     52013     60022: contig of 8010 bp in length
*     60023     60122: gap of 100 bp
*     60123     70438: contig of 10316 bp in length
*     70439     70538: gap of 100 bp
*     70539     77945: contig of 7407 bp in length
*     77946     78045: gap of 100 bp
*     78046     85717: contig of 7672 bp in length
*     85718     85817: gap of 100 bp
*     85818     92399: contig of 6582 bp in length
*     92400     92499: gap of 100 bp
*     92500     98215: contig of 5716 bp in length
*     98216     98315: gap of 100 bp
*     98316    104148: contig of 5833 bp in length
*    104149    104248: gap of 100 bp
*    104249    109166: contig of 4918 bp in length
*    109167    109266: gap of 100 bp
*    109267    112457: contig of 3191 bp in length
*    112458    112557: gap of 100 bp
*    112558    115076: contig of 2519 bp in length
*    115077    115176: gap of 100 bp
*    115177    117063: contig of 1887 bp in length
*    117064    117163: gap of 100 bp
*    117164    118652: contig of 1489 bp in length
*    118653    118752: gap of 100 bp
*    118753    120425: contig of 1673 bp in length
*    120426    120525: gap of 100 bp
*    120526    121843: contig of 1318 bp in length
*    121844    121943: gap of 100 bp
*    121944    123192: contig of 1249 bp in length.
FEATURES             Location/Qualifiers
     source          1. .123192
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q13"
                     /clone="CMB9-105N5"
     misc_feature    1. .15827
                     /note="assembly_fragment"
     misc_feature    15928. .30484
                     /note="assembly_fragment"
     misc_feature    30585. .40602
                     /note="assembly_fragment"
     misc_feature    40703. .51912
                     /note="assembly_fragment"
     misc_feature    52013. .60022
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    60123. .70438
                     /note="assembly_fragment"
     misc_feature    70539. .77945
                     /note="assembly_fragment"
     misc_feature    78046. .85717
                     /note="assembly_fragment"
     misc_feature    85818. .92399
                     /note="assembly_fragment"
     misc_feature    92500. .98215
                     /note="assembly_fragment"
     misc_feature    98316. .104148
                     /note="assembly_fragment"
     misc_feature    104249. .109166
                     /note="assembly_fragment"
     misc_feature    109267. .112457
                     /note="assembly_fragment"
     misc_feature    112558. .115076
                     /note="assembly_fragment"
     misc_feature    115177. .117063
                     /note="assembly_fragment"
     misc_feature    117164. .118652
                     /note="assembly_fragment"
     misc_feature    118753. .120425
                     /note="assembly_fragment"
     misc_feature    120526. .121843
                     /note="assembly_fragment"
     misc_feature    121944. .123192
                     /note="assembly_fragment"
ORIGIN

  Query Match            18.4%;  Score 460.6;  DB 14;  Length 123192;
  Best Local Similarity  59.8%;  Pred. No. 3.5e-83;
  Matches  882;  Conservative    1;  Mismatches  575;  Indels   17;  Gaps    6;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 656
              |||   ||||||||| || || ||| ||||||| || |  ||| |||  ||||||| ||
Db      63120 TTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 63179

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
              |  || |||  ||| |   | ||||  || ||||  ||||| | | |||||||  |   |
Db      63180 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 63239

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
              |||||||||||| |||||||| || |||  || ||| ||||||||||  ||| |||
Db      63240 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 63299

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
              |   || || |||||||| |||| |  | | ||||| |||||| ||||  | ||||||
Db      63300 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 63356

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
              | ||| |  |||| || |||| ||| ||||||||||||||||| ||| ||||||||
Db      63357 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 63416

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
              || |  |||| |  || |  |||  |    ||  ||    || |           |
Db      63417 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 63476

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
              |      ||     || |||| ||       | ||           |   ||  |   |
```

```
Db      63477 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 63536

Qy       1017 TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 1076

Db      63537 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCT 63596

Qy       1077 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--G 1134

Db      63597 CAGCCTCCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGT 63656

Qy       1135 TATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCT 1194

Db      63657 ATTTTTTAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 63716

Qy       1195 CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTG 1254

Db      63717 CAGGTGATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCG 63776

Qy       1255 CCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314

Db      63777 CCCGATGCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TA 63835

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374

Db      63836 TCATTTCCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATG 63895

Qy       1375 TTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434

Db      63896 ATTATTATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATT 63955

Qy       1435 TTTTCA---------CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTAT 1485

Db      63956 ATATCAATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAA 64015

Qy       1486 TGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTA 1545

Db      64016 TGGGTGTTTAGGTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGT 64075

Qy       1546 TACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACC 1605

Db      64076 TTCTAGAACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGT 64135

Qy       1606 CTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGT-TTTTTTTTAATGTT 1664

Db      64136 CCATGGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGC 64195

Qy       1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724

Db      64196 CAAAAATGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTTAAAATGATGTTGA 64255

Qy       1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784

Db      64256 AGGCCAGGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGG 64315

Qy       1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844

Db      64316 ATTACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTAC 64375

Qy       1845 TAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 1904

Db      64376 TAAAAATACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGA 64435

Qy       1905 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 1964

Db      64436 GGCTGAGGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGT 64495

Qy       1965 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATA 2024

Db      64496 GTCACTGCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATA 64555

Qy       2025 AAATAAAGGATCGGAGAGAAACAAAACTAATAAGA 2059

Db      64556 AAAATAATGAAAATAATAAAATGATGTTAAAGATA 64590
```

```
RESULT 9
AC022408
LOCUS       AC022408               150266 bp    DNA     linear   PRI 21-AUG-2001
DEFINITION  Homo sapiens chromosome 15 clone RP11-118J16 map 15q15, complete
            sequence.
ACCESSION   AC022408
VERSION     AC022408.6  GI:15217198
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Sequencing of human chromosome 15 D15S146-D15S117 region
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B.,
            Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G.,
            James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T.
            and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-FEB-2000) Multimegabase Sequencing Center, University
            of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE   3  (bases 1 to 150266)
  AUTHORS   Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S.,
            Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R.,
            Pate,D. and Hood,L.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-AUG-2001) Multimegabase Sequencing Center, Institute
            for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA
            98105, USA
COMMENT     On Aug 21, 2001 this sequence version replaced gi:11225361.
            ------------ Genome Center
                Center: Multimegabase Sequencing Center
                Center code: UWMSC
                Web site: http://chroma.mbt.washington.edu/msg_www
                Contact: leerowen@systemsbiology.org
            ------------ Summary Statistics
                Sequencing vector: pUC18; L08752
                Chemistry: Dye-terminator Big Dye; 90% of reads
                Chemistry: Dye-primer Big Dye; 10% of reads
                Assembly program: Phrap; version 0.990399
                Note:data from AC021753 [Drafting center UWMSC] and AC012652
            [Drafting center UWMSC] and 580I7 [Drafting center WUGSC] were
            added for finishing.
FEATURES             Location/Qualifiers
     source          1. .150266
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-118J16"
                     /clone_lib="RPCI human BAC library 11"
                     /note="This clone overlaps RP11-46M12 AC012652 and
                     RP11-129I12 AC021753 and RP11-580I7 AC068981.Data from
                     overlapping BACs were combined and the consensus sequence
                     determined from RP11-118J16."
     misc_feature    1. .150266
                     /note="overlap with RP11-580I7 AC068981."
     misc_feature    1. .100397
                     /note="overlap with RP11-129I12 AC021753."
     unsure          49355. .49359
                     /note="low quality data."
     misc_feature    100628. .150266
                     /note="overlap with RP11-46M12 AC012652."
ORIGIN
```

Query Match 18.4%; Score 460.4; DB 8; Length 150266;
Best Local Similarity 62.0%; Pred. No. 3.7e-83;
Matches 889; Conservative 1; Mismatches 487; Indels 57; Gaps 8;

```
Qy       600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
Db     52138 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 52197

Qy       660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
Db     52198 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 52257

Qy       720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
Db     52258 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 52317

Qy       780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
Db     52318 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 52375

Qy       840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
Db     52376 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCAA 52434

Qy       900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
Db     52435 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 52494

Qy       960 GGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
Db     52495 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 52553

Qy      1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
Db     52554 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 52613

Qy      1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
Db     52614 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 52673

Qy      1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
Db     52674 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 52733

Qy      1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
Db     52734 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 52793

Qy      1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
Db     52794 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 52845

Qy      1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
Db     52846 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 52891

Qy      1380 TTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
Db     52892 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 52930

Qy      1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
Db     52931 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 52990

Qy      1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
Db     52991 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 53050

Qy      1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
Db     53051 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 53105

Qy      1620 ATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
Db     53106 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 53165

Qy      1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
Db     53166 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 53225

Qy      1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
Db     53226 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 53283

Qy      1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
Db     53284 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 53343

Qy      1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
Db     53344 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 53403

Qy      1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
Db     53404 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACACCACTATACTC 53463

Qy      1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db     53464 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAACATTCA 53517
```

RESULT 10
AC021753

LOCUS	AC021753 167996 bp DNA linear PRI 29-OCT-2001
DEFINITION	Homo sapiens chromosome 15 clone RP11-129I12 map 15q15, complete sequence.
ACCESSION	AC021753
VERSION	AC021753.7 GI:16507213
KEYWORDS	HTG.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE	1 (bases 1 to 167996)
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE	Sequencing of human chromosome 15 D15S146-D15S117 region
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 167996)
AUTHORS	Rowen,L., Madan,A., Qin,S., Abbasi,N., Baradarani,L., Birditt,B., Bloom,S., Dors,M., Dickhoff,R., Fleetwood,P., Harrison,G., James,R., Kaur,A., Madan,A., Owen,M.P., Ratcliffe,A., Shaffer,T. and Hood,L.
TITLE	Direct Submission
JOURNAL	Submitted (20-JAN-2000) Multimegabase Sequencing Center, University of Washington, PO BOX 357730, Seattle, WA 98195, USA
REFERENCE	3 (bases 1 to 167996)
AUTHORS	Rowen,L., Madan,A., Qin,S., Baradarani,L., Birditt,B., Bloom,S., Burke,J., Dors,M., Fleetwood,P., Kaur,A., Madan,A., Nesbitt,R., Pate,D. and Hood,L.
TITLE	Direct Submission
JOURNAL	Submitted (29-OCT-2001) Multimegabase Sequencing Center, Institute for Systems Biology, 4225 Roosevelt Way NE, Suite 200, Seattle, WA 98105, USA
COMMENT	On Oct 29, 2001 this sequence version replaced gi:11136698. ---------- Genome Center Center: Multimegabase Sequencing Center Center code: UWMSC Web site: http://chroma.mbt.washington.edu/msg_www Contact: leerowen@systemsbiology.org ---------- Summary Statistics Sequencing vector: pUC18; L08752 Chemistry: Dye-terminator Big Dye; 90% of reads

```
            Chemistry: Dye-primer Big Dye; 10% of reads
            Assembly program: Phrap; version 0.990399
            Note: data from AC022408 [Drafting center UWMSC] and AC020661
        were added for fimishing.
FEATURES             Location/Qualifiers
     source          1. .167996
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="15"
                     /map="15q15"
                     /clone="RP11-129I12"
                     /clone_lib="RPCI Human BAC library 11"
                     /note="This clone overlaps RP11-118J16 AC022408 and
                     RP11-540O11 AC020661. Data from overlapping BACs were
                     combined and the consensus sequence determined from
                     RP11-129I12 to the extent possible."
     misc_feature    1. .63402
                     /note="overlap with RP11-540O11 AC020661."
     misc_feature    67596. .167996
                     /note="overlap with RP11-118J16 AC022408."
ORIGIN

  Query Match             18.4%;  Score 459.4;  DB 8;  Length 167996;
  Best Local Similarity   62.0%;  Pred. No. 5.8e-83;
  Matches  888;  Conservative    1;  Mismatches  487;  Indels   57;  Gaps    8;

Qy        600 TGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCA 659
Db     119755 TGTGTATTTTGTTGTTGTTGTTGTTGAGCCAGCTCGGGCTCTGTCGCCCAGGCTGGAGTG 119814

Qy        660 CAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGAC 719
Db     119815 CAGTGGCTCGATCTCGGCTCACTGCAATCTCTGCCTCCTGGGCTCAAGCCATCCTCTCAC 119874

Qy        720 CTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAA 779
Db     119875 CTCAGCATCCCAAGTAGCTGGGACTACAGGTGCACATCACCATGCTTAGCTAATTTTTGT 119934

Qy        780 AATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCT 839
Db     119935 A--TTTTCTGTAGAGACAGGGTTTTGCCATGTTGCCCATGCTGATTTCAAACTCCTGAGC 119992

Qy        840 TCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCAT 899
Db     119993 TCAAACGATCCGCCT-GCTCAGCCTCCCAAAGTGCCGGGATTACAGGTGTGAGCCACCAA 120051

Qy        900 GTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACA 959
Db     120052 GCCTAACCTATTTAATGTGAATTTGTATGTTGGCATATAAAATCCCCTAGCCAGTTTTGT 120111

Qy        960 GGGATTACTGCCTATTTCTTCCTTTCTTTTTGAGACAGAGTCTCACTTCATCACCTCA 1019
Db     120112 TTTGTT-TTGTTTTTTTGAGACAGAGTTTTGCTCTTGTTGCACTCCACTGGAGTGCAGTG 120170

Qy       1020 ACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAG 1079
Db     120171 GCTCAATCTCGGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCGATTCTCCTGCCTCAG 120230

Qy       1080 CCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTT 1139
Db     120231 CCTCCTGAGTAGCTGGGATTACAGGTGCGCGCCACCACGCCCAGCTAATTTTTTGTATTT 120290

Qy       1140 TTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGT 1199
Db     120291 TTGGTAGAGTCAGAGTTTAACCATGTTGGCCAGGCTGGTTTCGAACTCCTGACCTCAGGT 120350

Qy       1200 GATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGG 1259
Db     120351 GATCCACCTGCCTCAGTCTCCCAAAGTGCAGGGATTACAGGCGTGAGCCACCGCACCCGG 120410

Qy       1260 CCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATA 1319
Db     120411 CCAACAAACTTTCATTAAGTGAACCTCTAACACCACAGCTTATAATTAGTTT-------- 120462

Qy       1320 TAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCT 1379
Db     120463 --------------TGTTTTGTTTTTAGAGACAGGGTCTCGCTCTCCACCCAGGCTTGAG 120508

Qy       1380 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTC 1439
Db     120509 TGCGCTGGCACAATCATAGCTCACTAGGGACTCAA--------------------GTGA 120547

Qy       1440 ACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTT 1499
Db     120548 TCCTCTTGCCTCCTGAGTAACTGGGACTACAGGCTTGAGACACCATGCCCAGCAGAATAG 120607

Qy       1500 CATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCT 1559
Db     120608 CTCTCGATTTGTTGCTTTGGCTTAATCACATTTCCACCTTGATGTACTCAAAATGTTTTT 120667

Qy       1560 CCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAA 1619
Db     120668 ATTAGTCATA-----AACTTTTCTAATCACATTTGTGCTCACTACAATTATTTCCTATCA 120722

Qy       1620 ATACACACACACACACACACTCACACACAGTTTTTTTTAATGTTTGCAACTAAGACAAG 1679
Db     120723 CAATTTACAGACTTTCAAAAATTGATTTGGTTTCTACTGAAATAAAACGTAGAGGAAAAG 120782

Qy       1680 AAACCTGCA---TTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACA 1736
Db     120783 AACACAGTAAGCTAAAATGATGGAATTCAATCCCTTAAAAACAAAATTCAGCCGGGCACG 120842

Qy       1737 GTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGT 1796
Db     120843 GTGGCTCACGCCTGTAATCTCAGCACTTTGGGAGGCCAAGGTAGGTGGATCAC--GAGGT 120900

Qy       1797 GAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAA 1856
Db     120901 CAGGGGCTCGAGACCAGCCTGACCAACATGGTGAAACCCCGTCTCTACTAAAAATACAAA 120960

Qy       1857 AATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAG 1916
Db     120961 AATTAGCTGGGTGTGGTGGCGGGCGCCTGTAATACCAGCTACTCAGGAGGCTGAGGCAGG 121020

Qy       1917 AGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTC 1976
Db     121021 AGAATTGCTTGAACCCGGGAGGCGGAAGCTGCAGTGAGCCGAGATCACGCCACTATACTC 121080

Qy       1977 CAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATA 2029
Db     121081 CAGCCTGGGCGACAGAGCCAGACTCCGTCTCAAAAAACAAAAAAAAAAAACA 121133


RESULT 11
AC003689
LOCUS       AC003689              137693 bp    DNA     linear   PRI 21-NOV-1998
DEFINITION  Homo sapiens Chromosome 11q12.2 PAC clone pDJ1081b4 containing
            human mRNA for T-cell glycoprotein CD6, complete sequence.
ACCESSION   AC003689
VERSION     AC003689.1  GI:3900834
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 137693)
  AUTHORS   Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M.,
            Buettner,J., Bumeister,R., Card,P., deSailboat,F., Dunn,J.,
            English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G.,
            Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N.,
            Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S.,
            Schageman,J., Schultz,R.A., Stimson,S., Syed,M. and Ward,T.
  TITLE     HTGS Submission
  JOURNAL   Unpublished
```

REFERENCE 2 (bases 1 to 137693)
AUTHORS Evans,G.A., Athanasiou,M., Basit,M., Bradbury,P., Brignac,S., Bumesiter,R., Davis,C., English,C., Franklin,T.L., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J., Hinson,S., Narayanaswamy,U., Newton,J., O'Brien,K., Patel,P., Schageman,J., Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T. and Wilson,R.
TITLE Direct Submission
JOURNAL Submitted (17-DEC-1997) Genome Science & Technology Center, University of Texas Southwester Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
REFERENCE 3 (bases 1 to 137693)
AUTHORS Evans,G.A., Athanasiou,M., Aguayo,P., Armstrong,D., Basit,M., Buettner,J., Butler,C., Card,P., deSailboat,F., Dunn,J., English,C., Ethridge,S., Garner,H.R., Gee,V., Gordon,M., Gotway,G., Grant,O., Hahner,L., Joslin,J., Lewis,E., Loo,H., Loo,K.N., Major,T., McFarland,J., Newton,J., Osborne-Lawrence,S., Schageman,J., Schultz,R.A., Stimson,S., Waller,K. and Ward,T.
TITLE Direct Submission
JOURNAL Submitted (21-NOV-1998) Genome Science & Technology Center, University of Texas Southwestern Medical Center, 5323 Harry Hines Blvd, Dallas, TX 75235-8591, USA
COMMENT On Nov 21, 1998 this sequence version replaced gi:2695567.
Further information regarding the map of this region or annotation of pDJ1081b4 can be found at http://gestec.swmed.edu/chromoso1.htm
IMPORTANT: This submission contains the entire insert of clone pDJ1081b4. pDJ1081b4 comes from the RPCI-3 PAC library constructed at the Roswell Park Cancer Institute by the Pieter de Jong group. This clone has been finished according to strict quality criteria and attempts have been made to resolve all base calling problems such as compressions and repetitive elements. The expected Phred/Phrap calculated errors/10kb is 0.98. In addition, attempts have been made to assure over 99% of consensus base calls consist of either double-stranded coverage or 2 types of labeling chemistry on one strand.
CHROMOSOMAL LOCUS: This PAC clone comes from the Chromosome 11p12.2 Best's disease region mapped between STS D11S461 and EST AHNAK. This region spans over 1.5 Mbp.
MARKER CONFIRMATION: ESTs/STSs sequence confirmed; CD6, D11S461
MAPPED CLONE OVERLAP: PACs pDJ606g6 and pDJ59L21.

FEATURES Location/Qualifiers
source 1. .137693
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
repeat_region 164. .454
/rpt_family="Alu"
repeat_region complement(1543. .1835)
/rpt_family="Alu"
repeat_region complement(2058. .2360)
/rpt_family="Alu"
repeat_region complement(3515. .3796)
/rpt_family="Alu"
repeat_region complement(5954. .6244)
/rpt_family="Alu"
repeat_region complement(6962. .7215)
/rpt_family="Alu"
repeat_region complement(8088. .8263)
/rpt_family="Alu"
repeat_region complement(12551. .12719)
/rpt_family="Alu"
repeat_region 13480. .13676
/rpt_family="MER33"
repeat_region 13677. .13985
/rpt_family="Alu"
repeat_region complement(15815. .15950)
/rpt_family="Alu"
repeat_region complement(17641. .17934)
/rpt_family="Alu"
repeat_region 18423. .18693
/rpt_family="Alu"
repeat_region 18722. .19008
/rpt_family="Alu"
repeat_region complement(21315. .21594)
/rpt_family="Alu"
repeat_region complement(21636. .22481)
/rpt_family="MER42"
repeat_region complement(23396. .23697)
/rpt_family="Alu"
repeat_region 25083. .25349
/rpt_family="Alu"
repeat_region complement(28459. .28738)
/rpt_family="Alu"
repeat_region 29031. .29103
/rpt_family="Alu"
repeat_region complement(29491. .29785)
/rpt_family="Alu"
repeat_region complement(30052. .30172)
/rpt_family="Alu"
repeat_region complement(30261. .30640)
/rpt_family="Alu"
repeat_region 30885. .31193
/rpt_family="Alu"
repeat_region 32498. .32814
/rpt_family="Alu"
repeat_region complement(33478. .33782)
/rpt_family="Alu"
repeat_region 34395. .34556
/rpt_family="MIR"
repeat_region 35452. .35745
/rpt_family="Alu"
repeat_region complement(35739. .35983)
/rpt_family="MER33"
repeat_region complement(36141. .36244)
/rpt_family="MER3"
repeat_region complement(36206. .36244)
/rpt_family="MER33"
repeat_region 37244. .37491
/rpt_family="Alu"
repeat_region 38123. .38398
/rpt_family="Alu"
repeat_region complement(39547. .39837)
/rpt_family="Alu"
repeat_region 40124. .40270
/rpt_family="MER45"
repeat_region 40625. .40913
/rpt_family="Alu"
repeat_region complement(41015. .41075)
/rpt_family="MLT1"
repeat_region complement(41333. .41615)
/rpt_family="Alu"
repeat_region 41812. .42105
/rpt_family="Alu"
repeat_region 44026. .44309
/rpt_family="Alu"
repeat_region complement(47659. .47752)
/rpt_family="MER5"
repeat_region 50186. .50330
/rpt_family="MIR"
repeat_region 52843. .53128
/rpt_family="Alu"
repeat_region 53437. .53590
/rpt_family="MIR"
repeat_region 53688. .53967
/rpt_family="Alu"
repeat_region 56946. .57144
/rpt_family="Alu"
repeat_region 59439. .59731
/rpt_family="Alu"
repeat_region complement(61583. .61649)
/rpt_family="MIR"
repeat_region complement(62025. .62159)
/rpt_family="MIR"
repeat_region complement(62743. .63037)
/rpt_family="Alu"

```
     repeat_region   complement(63993. .64088)
                     /rpt_family="MIR"
     repeat_region   complement(66479. .66870)
                     /rpt_family="Alu"
     repeat_region   67416. .67584
                     /rpt_family="MER42"
     repeat_region   67594. .67887
                     /rpt_family="Alu"
     repeat_region   68119. .68182
                     /rpt_family="MER42"
     repeat_region   complement(69234. .69519)
                     /rpt_family="Alu"
     repeat_region   complement(69560. .69821)
                     /rpt_family="Alu"
     repeat_region   complement(70420. .70676)
                     /rpt_family="Alu"
     repeat_region   complement(70696. .70906)
                     /rpt_family="MER42"
     repeat_region   71247. .71537
                     /rpt_family="Alu"
     repeat_region   complement(71520. .71655)
                     /rpt_family="MER42"
     repeat_region   72778. .73053
                     /rpt_family="Alu"
     repeat_region   complement(75074. .75330)
                     /rpt_family="Alu"
     repeat_region   complement(76584. .76879)
                     /rpt_family="Alu"
     repeat_region   complement(76944. .77232)
                     /rpt_family="Alu"
     repeat_region   77719. .77994
                     /rpt_family="Alu"
     repeat_region   78902. .79179
                     /rpt_family="Alu"

  Query Match             18.3%;  Score 457.4;  DB 8;  Length 137693;
  Best Local Similarity   59.7%;  Pred. No. 1.5e-82;
  Matches  880;  Conservative    1;  Mismatches  577;  Indels   17;  Gaps    6;

Qy        598 TTTGGGTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGA 656
Db      76572 TTTCTTTCTTTCTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGA 76631

Qy        657 GCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCT 716
Db      76632 GTGCAATGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCC 76691

Qy        717 GACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAA 776
Db      76692 TGCCTCAGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTT 76751

Qy        777 AAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTG 836
Db      76752 GTA---TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTG 76808

Qy        837 GCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCAC 896
Db      76809 ACCTCAGGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCAT 76868

Qy        897 CATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGA 956
Db      76869 CACGCCCGGCCTCACATATTCTTATCCTAATATGTTTTCATGCTTTGAAGTCCTACAGTA 76928

Qy        957 ACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC 1016
Db      76929 ATCATCCTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCA 76988

Qy       1017 TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCT 1076
Db      76989 ATGGCGCAATCTCGGCTCACTGCAAACTCTGCCTCCCGAGGTTCACGCGATTCTCCTGCCT 77048

Qy       1077 AAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTT--G 1134
Db      77049 CAGCCTCCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTGT 77108

Qy       1135 TATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCT 1194
Db      77109 ATTTTTAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCT 77168

Qy       1195 CAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTG 1254
Db      77169 CAGGTGATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCG 77228

Qy       1255 CCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTG 1314
Db      77229 CCCGATGCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TA 77287

Qy       1315 ACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTT 1374
Db      77288 TCATTTCCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATG 77347

Qy       1375 TTTCTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCT 1434
Db      77348 ATTATTATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATT 77407

Qy       1435 TTTTCA---------CTTATCACCTTGTTATGACTTTTCATATTGCCTCAAACCTTTAT 1485
Db      77408 ATATCAATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAA 77467

Qy       1486 TGTTACTGTTTTTCATTGTTACTATTTAGTCACTGAATAATATGGCTTAATTTGCTTA 1545
Db      77468 TGGGTGTTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGT 77527

Qy       1546 TACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACC 1605
Db      77528 TTCTAGAACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGT 77587

Qy       1606 CTCTCCCCAGAGAAATACACACACACACACACTCACACACAGT-TTTTTTTAATGTT 1664
Db      77588 CCATGGAGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGC 77647

Qy       1665 TGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACT 1724
Db      77648 CAAAAATGTCACAATGTAATCTAGAGTCAAATTATTTTTTACTTTTAAAATGATGTTGA 77707

Qy       1725 CAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGG 1784
Db      77708 AGGCCAGGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGG 77767

Qy       1785 ATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTAC 1844
Db      77768 ATTACTTGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTAC 77827

Qy       1845 TAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGA 1904
Db      77828 TAAAAATACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGA 77887

Qy       1905 GGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCC 1964
Db      77888 GGCTGAGGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGT 77947

Qy       1965 ACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATA 2024
Db      77948 GTCACTGCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATA 78007

Qy       2025 AAATAAAGGATCGGAGAGAAACAAAACTAATAAGA 2059
Db      78008 AAAATAATGAAAATAATAAAATGATGTTAAAGATA 78042


RESULT 12
AP003721
LOCUS       AP003721               201460 bp    DNA     linear   PRI 17-JUN-2003
DEFINITION  Homo sapiens genomic DNA, chromosome 11 clone:RP11-881M11, complete
            sequence.
ACCESSION   AP003721
```

```
VERSION     AP003721.3  GI:31790667
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA
  JOURNAL   Published Only in Database (2001)
REFERENCE   2  (bases 1 to 201460)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (04-JUN-2001) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou,Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     On Jun 16, 2003 this sequence version replaced gi:28189463.
FEATURES             Location/Qualifiers
     source          1..201460
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11q"
                     /clone="RP11-881M11"
ORIGIN

  Query Match             18.3%;  Score 456.2;  DB 8;  Length 201460;
  Best Local Similarity   59.7%;  Pred. No. 2.5e-82;
  Matches  877;  Conservative    1;  Mismatches  574;  Indels   17;  Gaps    6;

Qy        604 TTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATGAGCACAG 662
Db     144460 TTTCTTTTTTTTTTTTTTTGAGACAGAGTTTCTCTCTTGTTGCCCAGGCTGGAGTGCAA 144519

Qy        663 TGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGATCTGCTGACCTC 722
Db     144520 TGGCACAATCTCGGTTCACCACAACCTCCGCCTCCCGGGTTCAAGGGTTTCTCCTGCCTC 144579

Qy        723 AGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAAT 782
Db     144580 AGCCTCCCGAGTAGCTGAGATTACAGGCATGCGCCACCACGCCCAGCTGATTTTGTA--- 144636

Qy        783 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCA 842
Db     144637 TTATTAGTAGAGACGGGGTTTCTCCATGTTGGTCAGGCTGGTCTCGAGCTCCTGACCTCA 144696

Qy        843 AGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTG 902
Db     144697 GGTGATCCACCCACCTCGGCCTCCCAAAGTGCTGGGATTATAGGTGTGAGCCATCACGCC 144756

Qy        903 CGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGG 962
Db     144757 CGGCCTCACATATTCTTATCCTAATATGTTTTCACGCTTTGAAGTCCTACAGTAATCATC 144816

Qy        963 ATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACC 1022
Db     144817 CTTTGATTCTTTTTTTTTGAGACAGTCTTGCTCTGTCCACCAGGCTGGAGTGCAATGGCG 144876

Qy       1023 TCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCT 1082
Db     144877 CAATCTCGGCTCACTGCAAACTCTGCCTCCGAGGTTCACGCGATTCTCCTGCCTCAGCCT 144936

Qy       1083 CCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT--GTATTTT 1140
Db     144937 CCAGAGTAGCTGGGATTACAGGTGCGCGCCACCACGTCTGGCTACTTTTTTTTGTATTTTT 144996

Qy       1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db     144997 TAGTAGAGACAGGGTTTCACCATGTTGGTCAGGCTGGTCTCGAACTCCTGACCTCAGGTG 145056

Qy       1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db     145057 ATCCGCCTGCCTCAGCCTTCCAAAGTGCTGGGATTACAGGCATGAGCCACCGCGCCCGAT 145116

Qy       1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db     145117 GCCCTTTGATTCTTTGAAACAAAATGTATATAAATTAAAATGCTAAAACGT-TATCATTT 145175

Qy       1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db     145176 CCTGTGATCATAAGGCTCTGACAACATTTTCCTCTAGATTATGTTGTCATTATGATTATT 145235

Qy       1381 TTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db     145236 ATTCACACATAAAGCAAAATAATATGGGTACTACAAATTTTGATGCTATTTATTATATCA 145295

Qy       1441 C---------TTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTATTGTTAC 1491
Db     145296 ATATAAAGGATGATTAGCCAGAGAAAACTAAATGTTATTGGAAAGGTATCTGAATGGGTG 145355

Qy       1492 TGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATC 1551
Db     145356 TTTAGGTTTTTCCCTCTCACTTAGTTAACATACCCAGAGATGGATATTATTTGTTTCTAG 145415

Qy       1552 CTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCC 1611
Db     145416 AACAAGACAGGAATTAGCATCCTGTTTTCATTTTTATAGATATGACCACTGAGTCCATGG 145475

Qy       1612 CCAGAGAAATACACACACACACACACACTCACACACAGT-TTTTTTTAATGTTTGCAAC 1670
Db     145476 AGCTAGAAGGAATTTGTTAGAGTGAATTTTCCCAACTCTGTCCAAGTTGGAGGCCAAAAA 145535

Qy       1671 TAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTG 1730
Db     145536 TGTCACAATGTAATCTAGAGTCAAATTATTTTTACTTTTAAAATGATGTTGAAGGCCA 145595

Qy       1731 GGCACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACT 1790
Db     145596 GGTGGGGTGGCTCACGTATGTAATCCAAGCACTTTGGAAGGCCGAGGCAGGTGGATTACT 145655

Qy       1791 TGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAA 1850
Db     145656 TGAGGTCAGGAGTTCGAGACCAGCCTGGCCAACATGGTGAAACACTGTCTCTACTAAAAA 145715

Qy       1851 TACAAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGA 1910
Db     145716 TACAAAAATTAGCTGGGCGTGGTGATGCGTGCCTGTAATCCCAGCTACTTGGGAGGCTGA 145775

Qy       1911 GGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACT 1970
Db     145776 GGCAGAAGAATCGCTTGAACCCGGGAGGCGGAGGTTGCAGTGAGCCGAGATTGTGTCACT 145835

Qy       1971 GCACTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAA 2030
Db     145836 GCACTCCAGCCTGGGCAACAGACCAAGATTCTGTCTCAAAATAAATAATAAATAAAAATA 145895

Qy       2031 AGGATCGGAGAGAAACAAAACTAATAAGA 2059
Db     145896 ATGAAAATAATAAAATGATGTTAAAGATA 145924

RESULT 13
AC116933
LOCUS       AC116933              195986 bp    DNA     linear   HTG 21-DEC-2002
DEFINITION  Papio anubis clone RP41-1C24, WORKING DRAFT SEQUENCE, 4 ordered
            pieces.
ACCESSION   AC116933
VERSION     AC116933.2  GI:27356688
KEYWORDS    HTG; HTGS_PHASE2; HTGS_DRAFT.
SOURCE      Papio anubis (olive baboon)
```

```
  ORGANISM  Papio anubis
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Cercopithecidae; Cercopithecinae; Papio.
REFERENCE   1  (bases 1 to 195986)
  AUTHORS   Akhter,N., Antonellis,A., Ayele,K., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Cariaga,K., Coleman,B., Engle,J., Granite,S., Guan,X., Gupta,J.,
            Haghighi,P., Han,J., Hansen,N., Ho,S.-L., Idol,J.R., Karlins,E.,
            Laric,P., Lee-Lin,S.-Q., Legaspi,R., Maduro,Q.L., Maduro,V.B.,
            Margulies,E.H., Masiello,C., Maskeri,B., McDowell,J.,
            Paguirigan,C., Pearson,R., Portnoy,M.E., Prasad,A.,
            Reddix-Dugue,N., Schandler,K., Schueler,M.G., Sison,C.,
            Stantripop,S., Thomas,J.W., Thomas,P.J., Touchman,J.W., Vogt,J.L.,
            Wetherby,K.D., Wiggins,L., Young,A. and Green,E.D.
  TITLE     NISC Comparative Sequencing Initiative
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (03-APR-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
REFERENCE   3  (bases 1 to 195986)
  AUTHORS   Green,E.D.
  TITLE     Direct Submission
  JOURNAL   Submitted (21-DEC-2002) NIH Intramural Sequencing Center, 8717
            Grovemont Circle, Gaithersburg, MD 20877, USA
COMMENT     On Dec 21, 2002 this sequence version replaced gi:19909407.
            ---------- Genome Center
                 Center: NIH Intramural Sequencing Center
                 Center code: NISC
                 Web site: http://www.nisc.nih.gov
                 Contact: nisc_zoo@nhgri.nih.gov
            ---------- Project Information
                 Center project name: cqw
                 Center clone name: 001C24

            The sequence data in this record represents an 'enhanced'
            version of a Phase 2 submission. Specifically, the indicated
            order and orientation of each sequence contig has been
            established using one or more of the following: read-pair
            data from individual subclones, overlaps with neighboring
            clones, alignment with available reference sequence (e.g.,
            human), and/or confirmation by PCR testing. In addition,
            the sequence assembly is based on at least 8X average
            coverage in Q20 bases and has been reviewed to rule out
            gross misassemblies, the low-quality ends of sequence
            contigs have been trimmed away, and each base is associated
            with a Phrap-derived quality score.
            ---------- Summary Statistics
                 Sequencing vector: plasmid; n/a; 100% of reads
                 Chemistry: Dye-terminator Big Dye; 100% of reads
                 Assembly program: Phrap; version 0.990319
                 Consensus quality: 195122 bases at least Q40
                 Consensus quality: 195370 bases at least Q30
                 Consensus quality: 195594 bases at least Q20
                 Insert size: 186000; agarose-fp
                 Insert size: 195686; sum-of-contigs
                 Quality coverage: 9.19x in Q20 bases; agarose-fp
                 Quality coverage: 8.74x in Q20 bases; sum-of-contigs
            ----------.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 4 contigs. Gaps between the contigs
            * are represented as runs of N. The order of the pieces
            * is believed to be correct as given, however the sizes
            * of the gaps between them are based on estimates that have
            * provided by the submittor.
            * This sequence will be replaced
            * by the finished sequence as soon as it is available and
            * the accession number will be preserved.
            *       1    76152: contig of 76152 bp in length
            *   76153    76252: gap of unknown length
            *   76253   126668: contig of 50416 bp in length
            *  126669   126768: gap of unknown length
            *  126769   177394: contig of 50626 bp in length
            *  177395   177494: gap of unknown length
            *  177495   195986: contig of 18492 bp in length.
FEATURES             Location/Qualifiers
     source          1. .195986
                     /organism="Papio anubis"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9555"
                     /clone="RP41-1C24"
                     /clone_lib="RP41"
     misc_feature    1. .76152
                     /note="assembly_fragment
                     clone_end:T7
                     vector_side:left"
     misc_feature    1. .54626
                     /note="clone overlaps with GenBank Accession Number
                     AC116934 clone RP41-206F5 (center project name cqv)"
     gap             76153. .76252
                     /estimated_length=unknown
     misc_feature    76253. .126668
                     /note="assembly_fragment"
     gap             126669. .126768
                     /estimated_length=unknown
     misc_feature    126769. .177394
                     /note="assembly_fragment"
     gap             177395. .177494
                     /estimated_length=unknown
     misc_feature    177495. .195986
                     /note="assembly_fragment
                     clone_end:SP6
                     vector_side:right"
ORIGIN

  Query Match             18.1%;  Score 451.2;  DB 14;  Length 195986;
  Best Local Similarity   61.1%;  Pred. No. 2.6e-81;
  Matches  891;  Conservative    1;  Mismatches  514;  Indels   52;  Gaps    8;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
              ||||  ||| || ||   ||||||||| ||||||||||||||||||||  |||  ||||
Db     180502 TTTTCCTTTTTTTTTAAGTAGAGACAGAGTCTTGCTCTGTCACCCAGGCTGGAGTGCAGT 180561

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
              ||||||   |   || ||||| ||| |  ||||| |  || | ||||  |  ||||||
Db     180562 AGTGCAATAGTGTTTCCCTGCATCCTTCAATTCCTGGGAACAGGTGATCCTCCCACCTCA 180621

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
              | || |  |||||||||||||    ||      |||||| |||  ||||| | ||||||||
Db     180622 GTCTTCTGAGTAGCTGGGACTTTAGGCACATGCCACCATGCCCAGCTAACTTAAAAAATT 180681

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
              |||||| ||| || |||||||| ||  ||||  ||||||  |||  ||||  | ||| ||||
Db     180682 TTTTTGAAGACACAGGGTCTTGCTCTGTTGCTCAGGCTACTCTCAAACAGCCGGCCTCAA 180741

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
              |  ||||||   |||   | |||||||||||||||||| ||| ||||||||||||  |
Db     180742 GTGATCCTCTCGCCTCATCCTCCCAAAGTGCTGGGATTAGAGGTGTGAGCCACCATGCCC 180801

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
              ||   || ||| || |  |   |  ||          | |||  |  |           |
Db     180802 AGCCCATTTTTTTTTAATTTCTTAATTTAATTTCTATTTTTAATTTTTA---TTATTTTA 180858

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
              ||| |     |||  ||  ||   |||||||||||||||||| | ||| |     |
Db     180859 TTATTTTTATTTTATTTATTTATATTTTTGAGACAGAGTCTCACTCCGTCGCACAGGCTG 180918

Qy       1024 CCGT-------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCT 1069
                ||             |||||||||||||||| ||||||||||||||: | ||
Db     180919 GAGTGCAATGGTGTGATCTCAGCTCACTGCAACCTCCGCCTCCCGGGTTCAAGCAACTCC 180978

Qy       1070 CCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATT 1129
              ||||||| |||||||||||||||||| |||||| | |  | |||| |  | ||||||| |
Db     180979 CCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGGCCTCCACTACACCTGGCTAA-T 181037
```

```
Qy       1130 TTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCT 1189
             ||||||||||||||| ||||| ||||||| ||||||||    |||||||||| |||||||
Db     181038 TTTTGTATTTTTAGTAGAGACGGGGTTTCACCATGTTAGTCAGGCTGGTCTTGAACTCCT 181097

Qy       1190 GACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCA 1249
             ||||||| |||| | |||||||| | ||||||||||||||| |||| ||||||||| ||
Db     181098 GACCTCAGGTGACCCGCCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCA 181157

Qy       1250 CTGTGCCTGGCCGATTACTGTCTATTTTCT----------TTATTGCTATATCCCCAGAT 1299
             ||||||| ||       | ||| |||| |          ||| |    ||  |  |||
Db     181158 CTGTGCCCAGCAACCCCATTTCTTTTTTTTAAACAAATGCTTACTAACCTACTCTAAGAG 181217

Qy       1300 CTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTA 1359
             ||  |   |   |||  |  ||  | | |   |||  |   |   |   |
Db     181218 CTGTAAGGGCAAATGAGGCAATGTCTTTACCCTCAGAATCTTATCTTTTGGTGAGGAGAT 181277

Qy       1360 GATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACA 1419
              ||| | |  |   |  |  |   | |||   ||| || |    ||   | || |
Db     181278 AATACACATATATAAATAATGATACAAGGACATATATGATAAATGCCATGACAGATTATA 181337

Qy       1420 ATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAAC 1479
             |      ||||        |      | ||    | |  ||    |||| ||| || |
Db     181338 AATCCATATTCAC-----TACACAAACAAGAGGCTTAGAAAAGTTCAGATTCCCACTTTG 181392

Qy       1480 CTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATT 1539
              |  || | |  |||       |        |||   |  |   |   ||
Db     181393 GGAGGCCGAGACAGGTGGATCACCTGAGGTCAGGAGTTCA--AGAGCAGCGTGGCCAACA 181450

Qy       1540 TGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTA 1599
             || | | ||  | ||       |   || ||      |           ||  |    |
Db     181451 TGGTGAAACTCCGTCTCTACTAAAAGTACAAAAATTAGCCGGGGGTAGTGGAGGGCACCT 181510

Qy       1600 TGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCACACACAGTTTTTTTTTA 1659
              || ||   |  | ||||            | |  | ||  || |||
Db     181511 GTAATCCCAGCTACTCAGAAGGCTGAGGCAGGAGAATCGCTTGAACCCAGGAGGCAGAGG 181570

Qy       1660 ATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGT---------------T 1704
              ||     | |  |||     |  ||||| |  ||  ||  |               |
Db     181571 CTGCAGTGAGCCGAGATTGCACCACTGCACTCTAGCCTGGGGGACAGAGCCAGACTCCAT 181630

Qy       1705 CATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTACTT 1764
             |  |  ||  ||||| ||    |  |||| ||||| ||||  ||||||| | ||| ||||
Db     181631 CTCAAAAAAAAAAAAAAAAAAGGCCGGGCGCAGTGGCTCACACCTGTAATCCCAGCACTT 181690

Qy       1765 TGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAATA 1824
             | | || |  ||| |||  |||||  |||||| || ||||||||||||||||| |||||
Db     181691 TAGGAGCCTGAGGCGGGCAGATCA--TGAGGTCAGGAGTTCGAGACCAGCCTGGCCAATA 181748

Qy       1825 TGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGCCT 1884
             ||||||||||  |||||||||||||||||||||||||||||| || |||    |  |||
Db     181749 TGGTGAAACCCCGTCTCTACTAAAAATACAAAAATTAGCTGGGCGTGGTGCCACGCACCT 181808

Qy       1885 GTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGAGG 1944
             ||||||||||||||| |||| |||||| |||||| | ||||||  |||||  ||||
Db     181809 GTAGTCCCAGCTACTCAAGAGGCCGAGGCAGGAGAATCACCTGAACCCAGGAGGTGGAGG 181868

Qy       1945 TTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCTAT 2004
             |||||||||||||| ||    ||| |||||||||||||| | || ||||||||| ||
Db     181869 TTGCAGTGAGCCGAAATGGTGCCATTGCACTCCAGCCTGGGTGTCAGAGCGAGACTCCAT 181928

Qy       2005 CTCAAAAAAATAAATAAA 2022
             |||||||| | ||| |||
Db     181929 CTCAAAAAGAAAAAGAAA 181946
```

```
RESULT 14
AC005052/c
LOCUS       AC005052              134210 bp    DNA     linear   PRI 08-NOV-2000
DEFINITION  Homo sapiens BAC clone CTB-38K21 from Xq23, complete sequence.
ACCESSION   AC005052
VERSION     AC005052.2  GI:10122134
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 134210)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 134210)
  AUTHORS   Tin-Wollam,A., Graves,T. and Cofman,M.
  TITLE     The sequence of Homo sapiens BAC clone CTB-38K21
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (12-JUN-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 134210)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-SEP-2000) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 134210)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (08-NOV-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Sep 14, 2000 this sequence version replaced gi:3212939.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_RG038K21
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.

            MAPPING INFORMATION:
            This sequence was generated from part of bacterial clone contigs of
            human chromosome X, constructed by the chromosome X mapping group
            at the Sanger Centre, Wellcome Trust Genome Campus, Hinxton, UK.
            Further information can be found at
            http://www.sanger.ac.uk/HGP/ChrX/

            SOURCE INFORMATION:
            Clone CTB-38K21 is from the first release of the human BAC library
            CITB-978SK-B.  The library contains cloned DNA from the male
            fibroblast cell line 978SK.  See:  Shizuya et al., Proc. Natl.
            Acad. Sci. USA 89:8794-7 (1992); U-J. Kim et al., Genomics 34:213-8
            (1996).  This clone is available from Research Genetics, Inc.
            (http://www.resgen.com).
            VECTOR:  pBeloBAC11
            Selection:  chloramphenicol

            NEIGHBORING SEQUENCE INFORMATION:
            The clone sequenced to the right is RP3-327A19, 200 base pair
```

```
          overlap.  Actual start of this clone is at base position 1 of
          CTB-38K21; actual end is at base position 9416 of RP3-327A19.
FEATURES             Location/Qualifiers
     source          1. .134210
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="X"
                     /map="Xq23"
                     /clone="CTB-38K21"
                     /clone_lib="CITB-978SK-B"
     repeat_region   20. .268
                     /rpt_family="Alu"
     repeat_region   269. .561
                     /rpt_family="Alu"
     repeat_region   978. .1285
                     /rpt_family="Alu"
     misc_feature    2964. .4303
                     /note="CpG_island (%GC=68.3, o/e=0.70, #CpGs=110)"
     misc_feature    3597. .3895
                     /note="similar to  EST BE299003 (NID:g9182751)"
     repeat_region   6613. .6919
                     /rpt_family="Alu"
     repeat_region   7154. .7214
                     /rpt_family="L2"
     repeat_region   7640. .7924
                     /rpt_family="Alu"
     repeat_region   8152. .8333
                     /rpt_family="MER1_type"
     repeat_region   8565. .8749
                     /rpt_family="MaLR"
     repeat_region   8750. .9044
                     /rpt_family="Alu"
     repeat_region   9045. .9256
                     /rpt_family="MaLR"
     repeat_region   9271. .9700
                     /rpt_family="L2"
     repeat_region   10238. .10290
                     /rpt_family="L2"
     repeat_region   10291. .10585
                     /rpt_family="Alu"
     repeat_region   10586. .10666
                     /rpt_family="L2"
     repeat_region   10667. .10822
                     /rpt_family="Alu"
     repeat_region   10823. .11015
                     /rpt_family="L2"
     repeat_region   11016. .11199
                     /rpt_family="MER1_type"
     repeat_region   11200. .11309
                     /rpt_family="L2"
     repeat_region   11310. .11544
                     /rpt_family="Alu"
     repeat_region   11545. .11610
                     /rpt_family="L2"
     repeat_region   11860. .12166
                     /rpt_family="Alu"
     misc_feature    12126. .12175
                     /note="similar to  EST AI025436 (NID:g3241049) ow27h07.s1"
     repeat_region   12178. .12228
                     /rpt_family="L1"
     repeat_region   12405. .12520
                     /rpt_family="L1"
     repeat_region   12712. .13008
                     /rpt_family="Alu"
     repeat_region   13009. .13315
                     /rpt_family="Alu"
     repeat_region   13603. .13850
                     /rpt_family="ERVL"
     repeat_region   13851. .14147
                     /rpt_family="Alu"
     repeat_region   14148. .14222
                     /rpt_family="ERVL"
     repeat_region   14223. .14526
                     /rpt_family="Alu"
     repeat_region   14527. .14713
                     /rpt_family="ERVL"
     repeat_region   14815. .15244
                     /rpt_family="L2"
     repeat_region   15245. .15566
                     /rpt_family="Alu"
     repeat_region   15567. .16446
                     /rpt_family="L2"
     repeat_region   16447. .16580
                     /rpt_family="Alu"
     repeat_region   16581. .16866
                     /rpt_family="L2"
     repeat_region   16867. .17178
                     /rpt_family="Alu"
     repeat_region   17179. .17251
                     /rpt_family="L2"
     repeat_region   17255. .17390
                     /rpt_family="MER1_type"
     repeat_region   17872. .18165
                     /rpt_family="Alu"
     repeat_region   18239. .18532
                     /rpt_family="Alu"
     repeat_region   18540. .18841
                     /rpt_family="Alu"
     misc_feature    18631. .18876
                     /note="similar to  EST AW867673 (NID:g8001829)"
     repeat_region   19286. .19570
                     /rpt_family="L2"
     repeat_region   19885. .20185
                     /rpt_family="Alu"
     repeat_region   21149. .21446
                     /rpt_family="Alu"
     repeat_region   21559. .21803
                     /rpt_family="Alu"
     repeat_region   21835. .21951
                     /rpt_family="L1"
     repeat_region   21979. .22286
                     /rpt_family="Alu"
     repeat_region   23260. .23565
                     /rpt_family="Alu"
     repeat_region   23568. .23864
                     /rpt_family="Alu"
     misc_feature    23875. .23924
                     /note="similar to  EST AI735654 (NID:g5057178) at19b08.x1"
     misc_feature    23875. .23924
                     /note="similar to  EST AW619972 (NID:g7326156)"
     misc_feature    23878. .23926

  Query Match           18.0%;  Score 449.2;  DB 8;  Length 134210;
  Best Local Similarity   59.5%;  Pred. No. 6.9e-81;
  Matches  896;  Conservative    1;  Mismatches  554;  Indels   55;  Gaps    6;

Qy       555 TGAGCTATTTGCCCAAGATCACACAGCTTGTAAGTGGTGACAGTTTGGGTTTTTTTTTGT 614
            || | |||||||   | |         | |   |||| |  |  ||    ||| ||| |
Db     53793 TGTGGTATTTGCTTTATTTATTTATTTATTTTTTGGTTAGGGGGTGCTGATTTCTTTTT 53734

Qy       615 TGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCAT 674
             | || ||| |||||||||||||| |||||||||||||||  |||  |||||   | | |||
Db     53733 TCTTTTTTTGAGACAGGGTCTTTCTCTGTCACCCAGGCTGGAGTGCAGTGACAC-ATCAT 53675

Qy       675 AGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGT 734
             | |||||||||||||||||||||| |||||||  |||  |  | ||||||||||| |||
Db     53674 GGCTCACTGCAGCCTCAACCTCCTGGGCTCAAGCAATCCTCAAATCTCAGCCTCCCGAGT 53615

Qy       735 AGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAG 794
             ||||||| |||   || ||  | |||| ||| |||||||||   ||| | ||||
Db     53614 AGCTGGGTCTATAGGCTTGAGCTACCAGGCCCGGCTAATTTTTAAATTGTTTTGTAGAGA 53555

Qy       795 ACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCT 854
             |||||| |||| ||   ||||| |||||||||||||||||| ||||||  ||||||||
Db     53554 CAAGGGTCTCACTATGTCACCCAGGGTTGTCTTAAACTCCTGGGTTCAAGGGATCCTCCT 53495
```

```
Qy       855 ACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATT 914
Db     53494 GCCTTGGCCTCCCAAAGTGCTAGGATTACAGGCATGAGCCACTGCACCCGGCCTGATTCC 53435

Qy       915 TCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTAT 974
Db     53434 TGATTTCTTTTTTCTTTTT----------------------------------TTCTTTCT 53409

Qy       975 TTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCA---- 1030
Db     53408 TTTTCTTTTTTTTTTTTTGAGACGGAGTCTCGCTCTGTCGCCCAGGCTGGAGTGCAGTGG 53349

Qy      1031 ----------GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGC 1080
Db     53348 CACGATCTTGGCTCACTGCAAGGTCCACCTCCTGGGTTCACGCCATTCTCCTGCCTCAGC 53289

Qy      1081 CTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTT 1140
Db     53288 CTCCGGAGTAGCTGGGACTACAGGTGCCCGCCACCACACCCAGCTAAATTTTTGTATTTT 53229

Qy      1141 TAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTG 1200
Db     53228 TAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCGATCTCCTGACCTC--GTG 53171

Qy      1201 ATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGC 1260
Db     53170 ATCCGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAAGCATGAGCCACCGCGCCTGGC 53111

Qy      1261 CGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATAT 1320
Db     53110 CTTTTTCTTTTTTATTTTTTTTGAGACAGGGCTTC-CCTCTGTCACCCAGGCTGTAATGC 53052

Qy      1321 AGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTT 1380
Db     53051 ACTGGCACTATCTTGGCTCACTGCAACCTCCACTTCCCAGGTTCAAGCAATTATCATGCC 52992

Qy      1381 TTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCA 1440
Db     52991 TCAGTCTCCTGAGTAACTGGGATTACAGGCGTGTGCCGCTACGCCCAGCTAATTTTTGTA 52932

Qy      1441 CTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAAC---CTTTATTGTTACTGTTTT 1497
Db     52931 TTTTTAGTAAAGAGGAGGTTTCCCCATGTTGGCCAAGATGGTCTCCATCTCTTGACCTCG 52872

Qy      1498 TTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTG 1557
Db     52871 TGATCTGCCCGCCTCGGCCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACTGTGCCCG 52812

Qy      1558 CTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAG 1617
Db     52811 GCCTGATAACAGCTTTCTATGTGCCAGCTACTGCTCCAAGTTTATATGTATTATATGTAT 52752

Qy      1618 AAATACACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACA 1677
Db     52751 TAATTCATTAGCTCATTTGTATTTATATGTACTATTAATTCCTTTTATTAGTCCATTTTC 52692

Qy      1678 AGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAG 1737
Db     52691 ACCCTGCTGATAAAGACATACTTGAGACTGAGAAGAAAAAGATGTTTAGGCCAGGAGTGG 52632

Qy      1738 TGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTG 1797
Db     52631 TAGCTCACACCTGTAATCCCAGCACTTTGGGAGGCCGAGGTGGAAGGATTACCTGAGATC 52572

Qy      1798 AGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAA 1857
Db     52571 AGGAGTTTGAGACCACCCTGGCCAACATGGCAAAACCCCATCTCTACTAAAAATACAAAA 52512

Qy      1858 ATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGA 1917
Db     52511 ATTAGTAGGGCATGGTGGTACATGCCTGTGATCTCAGCTACTTGGGAGGCTGAGGCAGGA 52452

Qy      1918 GAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCC 1977
Db.    52451 GAATTGCTTGAACCTGGGAGGTGGAGGTTGTAGTGAGCCAGGATCACGCCACTGCACTCC 52392

Qy      1978 AGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCG 2037
Db     52391 AGCCTGGGTGACAGAATGAGACTCCATCTAAAAAAAAAAAAAAAAAAGAAAGAAAAAGAA 52332

Qy      2038 GAGAGA 2043
Db     52331 AAAAGA 52326
```

RESULT 15
AC005484/c

```
LOCUS       AC005484              131943 bp    DNA     linear   PRI 30-SEP-2000
DEFINITION  Homo sapiens PAC clone RP5-847O8 from 14q24.3, complete sequence.
ACCESSION   AC005484
VERSION     AC005484.2  GI:5091654
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 131943)
  AUTHORS   Sulston,J.E. and Waterston,R.
  TITLE     Toward a complete human genome sequence
  JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
   PUBMED   9847074
REFERENCE   2  (bases 1 to 131943)
  AUTHORS   Cloud,J., Wohldmann,P. and Holmes,A.
  TITLE     The sequence of Homo sapiens PAC clone RP5-847O8
  JOURNAL   Unpublished
REFERENCE   3  (bases 1 to 131943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-AUG-1998) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   4  (bases 1 to 131943)
  AUTHORS   Waterston,R.H.
  TITLE     Direct Submission
  JOURNAL   Submitted (17-JUN-1999) Genome Sequencing Center, Washington
            University School of Medicine, 4444 Forest Park Parkway, St. Louis,
            MO 63108, USA
REFERENCE   5  (bases 1 to 131943)
  AUTHORS   Waterston,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-SEP-2000) Department of Genetics, Washington
            University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT     On Jun 17, 1999 this sequence version replaced gi:3907511.
            -------------- Genome Center
                Center: Washington University Genome Sequencing Center
                Center code: WUGSC
                Web site: http://genome.wustl.edu/gsc
                Contact: sapiens@watson.wustl.edu
            -------------- Summary Statistics
                Center project name: H_DJ0847O8
            --------------.

            NOTICE:  This sequence may not represent the entire insert of this
            clone.  It may be shorter because we only sequence overlapping
            clone sections once, or longer because we provide a small overlap
            between neighboring data submissions.

            This sequence was finished as follows unless otherwise noted:
            all regions were double stranded, sequenced with an alternate
            chemistry, or covered by high quality data (i.e., phred quality >=
            30); an attempt was made to resolve all sequencing problems, such
            as compressions and repeats; all regions were covered by sequence
            from more than one subclone; and the assembly was confirmed by
            restriction digest.
```

MAPPING INFORMATION:
This clone from chromosome 14 was provided by Dr. Pieter de Jong, Roswell Park Cancer Institute, Human Genetics Department, Elm and Carlton Streets, Buffalo NY 14263-0001 USA.

SOURCE INFORMATION:
This clone was derived from human PAC library RPCI-5, prepared by Pieter de Jong and coworkers at the Roswell Park Cancer Institute (http://bacpac.med.buffalo.edu) using the method described by Ioannou et al., Nature Genetics 6:84-9 (1994). The library is from one male donor.
The clone may be obtained either from Genome Systems, Inc. (http://www.genomesystems.com) or Research Genetics, Inc. (http://www.resgen.com); or from Pieter de Jong.
VECTOR: pCYPAC2
NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP5-892G5; the clone sequenced to the right is RP4-592G7, 200 bp overlap. Actual start of this clone is at base position 1 of RP5-847O8; actual end is at base position 131747 of RP5-847O8.

FEATURES	Location/Qualifiers
source	1. .131943
	/organism="Homo sapiens"
	/mol_type="genomic DNA"
	/db_xref="taxon:9606"
	/chromosome="14"
	/map="14q24.3"
	/clone="RP5-847O8"
	/clone_lib="RPCI-5"
repeat_region	1058. .1098
	/rpt_family="AT_rich"
repeat_region	1762. .2225
	/rpt_family="MaLR"
repeat_region	2305. .2587
	/rpt_family="Alu"
repeat_region	2937. .3086
	/rpt_family="L2"
repeat_region	3293. .3591
	/rpt_family="Alu"
repeat_region	3651. .3916
	/rpt_family="L1"
repeat_region	3917. .3943
	/rpt_family="(TA)n"
repeat_region	4815. .4841
	/rpt_family="AT_rich"
repeat_region	5153. .5244
	/rpt_family="L1"
repeat_region	5728. .6008
	/rpt_family="Alu"
repeat_region	6009. .6034
	/rpt_family="(CAA)n"
misc_feature	8080. .8468
	/note="match to EST R05271 (NID:g755891) ye91b03.s1"
STS	8080. .8277
repeat_region	8284. .8589
	/rpt_family="Alu"
misc_feature	8649. .9039
	/note="match to EST R05328 (NID:g755948) ye91b03.r1"
repeat_region	8857. .9007
	/rpt_family="MIR"
repeat_region	9062. .9152
	/rpt_family="MIR"
repeat_region	9216. .9288
	/rpt_family="MIR"
repeat_region	9476. .9770
	/rpt_family="Alu"
STS	9711. .10034
repeat_region	10004. .10114
	/rpt_family="Alu"
repeat_region	10115. .10135
	/rpt_family="(TAAA)n"
repeat_region	10136. .10158
	/rpt_family="Alu"
repeat_region	10205. .10501
	/rpt_family="Alu"
repeat_region	10513. .10819
	/rpt_family="Alu"
repeat_region	10890. .11242
	/rpt_family="MaLR"
repeat_region	11246. .11519
	/rpt_family="Alu"
repeat_region	11567. .11996
	/rpt_family="L2"
repeat_region	12006. .12113
	/rpt_family="L2"
repeat_region	12123. .12146
	/rpt_family="(CAAAA)n"
repeat_region	12147. .12430
	/rpt_family="Alu"
repeat_region	12445. .12535
	/rpt_family="L1"
repeat_region	12527. .13097
	/rpt_family="L1"
repeat_region	13217. .13339
	/rpt_family="Alu"
repeat_region	13389. .13994
	/rpt_family="L2"
repeat_region	14138. .14158
	/rpt_family="AT_rich"
repeat_region	14195. .15071
	/rpt_family="L1"
repeat_region	15048. .15521
	/rpt_family="L1"
repeat_region	15520. .15992
	/rpt_family="L1"
repeat_region	15989. .16221
	/rpt_family="L1"
repeat_region	16257. .16556
	/rpt_family="Alu"
repeat_region	16569. .16963
	/rpt_family="L1"
repeat_region	17016. .17039
	/rpt_family="AT_rich"
repeat_region	17073. .17182
	/rpt_family="MIR"
repeat_region	17252. .17452
	/rpt_family="MaLR"
repeat_region	17473. .17991
	/rpt_family="L2"
repeat_region	18106. .18192
	/rpt_family="MER1_type"
repeat_region	18193. .18504
	/rpt_family="Alu"
repeat_region	18505. .18588
	/rpt_family="MER1_type"
repeat_region	18624. .18904
	/rpt_family="L1"
repeat_region	19144. .19368
	/rpt_family="Alu"
repeat_region	20030. .20159
	/rpt_family="Alu"
repeat_region	20167. .20317
	/rpt_family="Alu"
repeat_region	20318. .20617
	/rpt_family="Alu"
repeat_region	20618. .20753
	/rpt_family="Alu"
repeat_region	20802. .20954
	/rpt_family="L1"
repeat_region	20955. .21256
	/rpt_family="Alu"

Query Match 17.9%; Score 447.8; DB 8; Length 131943;
Best Local Similarity 61.3%; Pred. No. 1.3e-80;
Matches 913; Conservative 1; Mismatches 523; Indels 52; Gaps 10;

```
Qy       597 GTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTC-TGTCACCCAGGCATG 655
Db    108973 GATTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTCGCTCTTGTCTCCCAGGCTGG 108914

Qy       656 AGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGC 715
Db    108913 AGTGCAGTGGCACGATCTCGGCTGACTGCAACTTCCGCCTCCCGGGTTCAAGCGATTCTC 108854

Qy       716 TGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTA 775
Db    108853 CTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGCACCTGCCACCACACCTGGCTAACTT 108794

Qy       776 AAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCT 835
Db    108793 TTGTA--TTTTTAGTAGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTTGAACTCCT 108736

Qy       836 GGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCA 895
Db    108735 GACCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACATGTGTGAGCTA 108676

Qy       896 CCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAG 955
Db    108675 CTGTGCCTGGCCTCAAACTTGATTTTTGCAATGTGATAGGTGAGAAATGGGTATTTTTCA 108616

Qy       956 AACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCAC 1015
Db    108615 ATTTACGTTTCTCTCTTTTTTTTTTTTTTCTTTTTTTGAGGCGGAGTCTTGCTCTGTCAC 108556

Qy      1016 CTCAACCTCCGT--------------TCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAA 1061
Db    108555 CCAGGCTGGAGTGCAATGGTGTGATCTCAGCTCACTGCAACCTCCACCTCCCAGGTTCAA 108496

Qy      1062 GYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTT 1121
Db    108495 TTGATTCTCCTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGTGCGTGCCACCATGCCT 108436

Qy      1122 GGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTC 1181
Db    108435 GGCTAA-TTTTTGTATTTTTAGTAGAAGTGGGGTTTCACAATGTTGGCCAGGCTGGTCTT 108377

Qy      1182 AAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGC 1241
Db    108376 GAACTCCTGACCTC--GTGATCCACCCGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGT 108319

Qy      1242 GTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTT-------------------CTTTA 1282
Db    108318 GTGAGCCACCACGTCTGGCCCAATTTACATTTCTCTTATTAGGAGCTAGATTGGACTTTT 108259

Qy      1283 TTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATT 1342
Db    108258 TTTTCACATTTGAAGGCCATTTGCATTTCTTTTAGTACCTTTTGTTCATTTGTTGTTATT 108199

Qy      1343 GATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAAC 1402
Db    108198 GGCCTTTCCTTCTTCCCCCCTTCCTTTTCTCTGTCCTCTTGCTTTCACTTCCCTCTCCTC 108139

Qy      1403 TTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTT 1462
Db    108138 CTCCTCTTTCTCCTTCTCTTCCTCTCAGTCTTTTATGTCTCCTTCCCTTTTTCTCCCTCC 108079

Qy      1463 TTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTG 1522
Db    108078 TCTTCCTTCCTCATCCTCCTTCTTTCTCCTACCTTCCTCTTTATTAGGGATATTTACCAC 108019

Qy      1523 AATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTAC 1582
Db    108018 ACTCCCTTTGACTCCTAGCCAATTAATTTTAACTAGTTTCGGGTATTAGGGTTATGTCAG 107959

Qy      1583 AAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTC- 1641
Db    107958 AGATAAAAGGTGTGTTAGGTCTTCTCATATCCCTTTTTAGTAAACATCTGCATTTTTTCC 107899

Qy      1642 -------ACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAG 1694
Db    107898 TTTCTATATAGAAACTTTTCCAAAATTTTTTTCTTATTCAATGCAGATGGTATCTTTTTA 107839

Qy      1695 GATGTTTGTTCATATTAATTAAAAATAACTCAGTT---GGGCACAGTGACTCAAGCCTGT 1751
Db    107838 CAGAAATGAAAACTTTATGCAAAAATAAAAGGGGGACCGGGCGCGGTGGCTCATGCCTGT 107779

Qy      1752 AACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACC 1811
Db    107778 AATCCCAGTACTTTGGGAGGCCAAAGCGGGTGGATCACCTGAGGTCAGGAGTTCAAGACC 107719

Qy      1812 AGCCTGGTCAATATGGTGAAACCCTATCTCTACT-AAAAATACAAAAATTAGCTGGGTGT 1870
Db    107718 AGCCTGGCCAACATCGTGAAACCCCATCTCTACTAAAAAATACAAAAATTAGCTGGGCGT 107659

Qy      1871 AGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCA-AGAGAATTGCTTGAA 1929
Db    107658 GGTGGCAGGCATCTGTAATCCCAGCTACGCGGGAGGCTGAGGCAGGGAGAATTGCTTGAA 107599

Qy      1930 CCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGAC 1989
Db    107598 CCTGGGAGGCAGAGGTTGCAGTGAGCTGAGATCACGCCACTACACTCCAGCCTGAGCAAC 107539

Qy      1990 ACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGGATCGG 2038
Db    107538 AGAGTGAGACTCCATCTCAAAAAAATAATAATAAATAAAGGCCAG 107490
```

Search completed: February 17, 2006, 13:51:33
Job time : 12648 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.7

OM nucleic - nucleic search, using sw model

Run on: February 17, 2006, 10:20:55 ; Search time 9872 Seconds
(without alignments)
11853.170 Million cell updates/sec

Title: US-10-607-806-1-C7256_COPY_7000_9500
Perfect score: 2499
Sequence: 1 gtcgtgtcactgctgtccag.....tttgagaccagcctggacaa 2501

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_htc:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_est7:*
9: gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

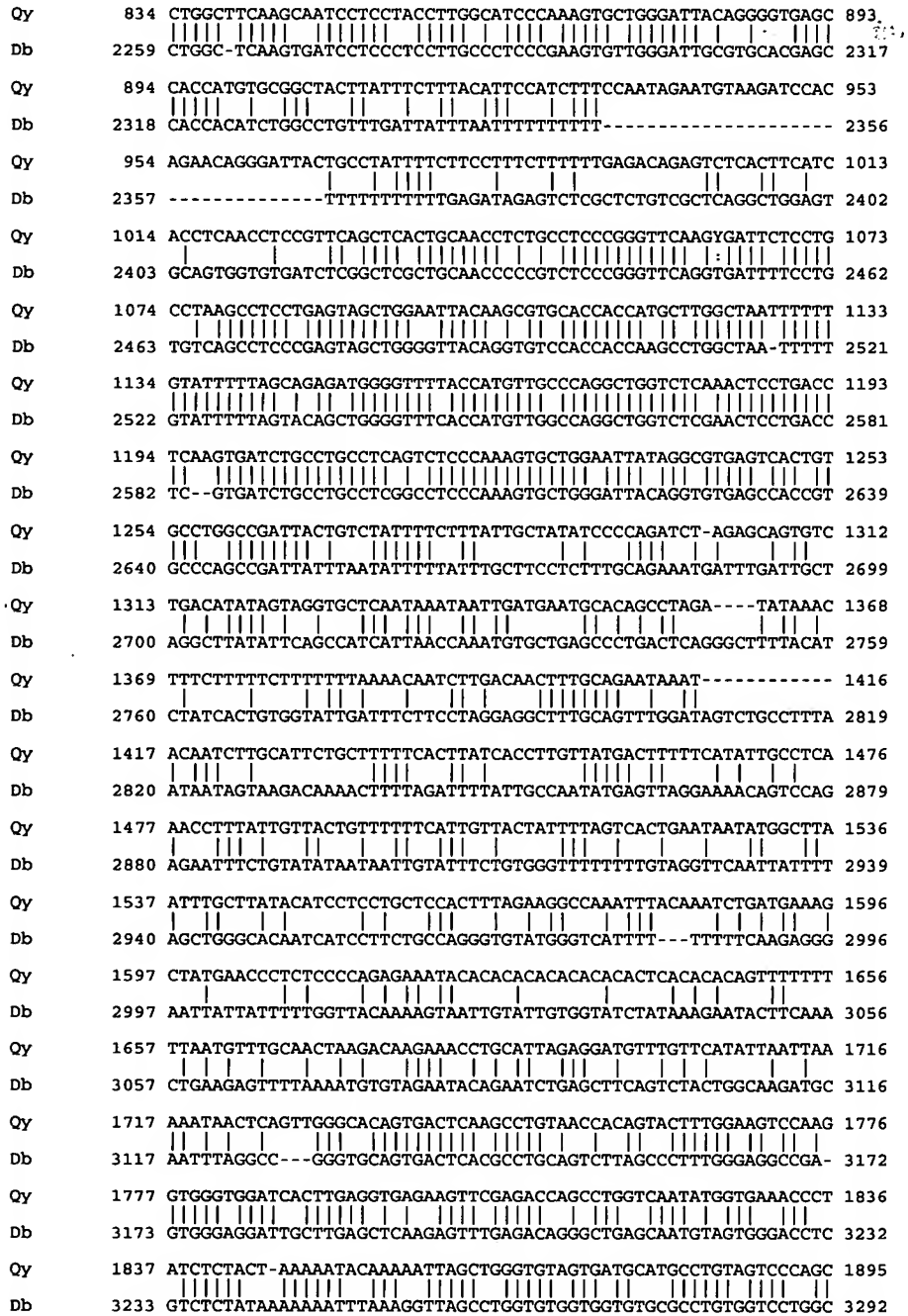
SUMMARIES

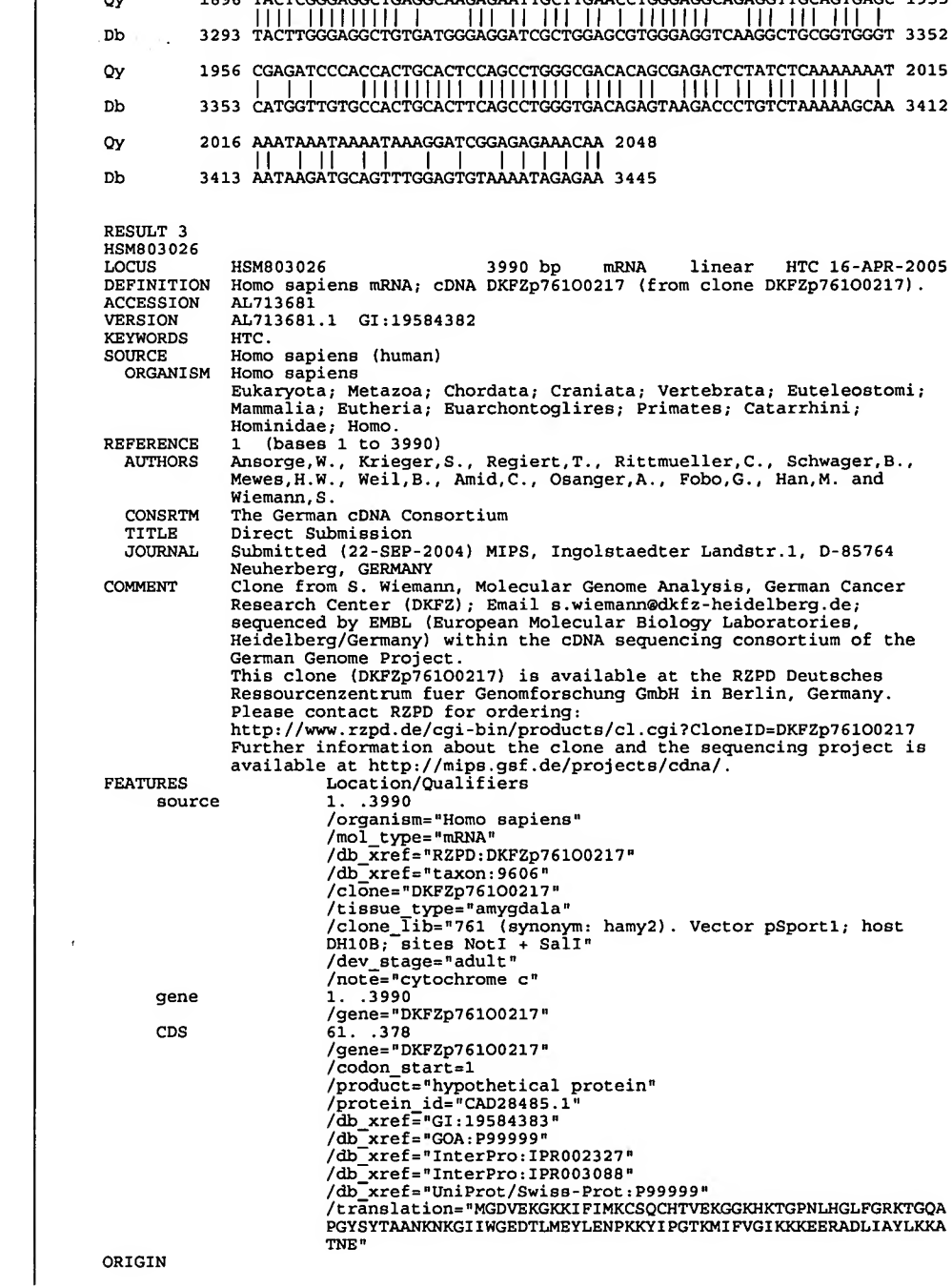
	Result No.	Score	% Query Match	Length	DB	ID	Description
c	1	505.4	20.2	507	5	BU785040	BU785040 in43g09.x
	2	344.2	13.8	4087	4	BC024593	BC024593 Homo sapi
	3	341	13.6	3990	4	HSM803026	AL713681 Homo sapi
c	4	337.2	13.5	15970	9	AQ839852	AQ839852 260L13-C5
c	5	315.8	12.6	897	6	CD557847	CD557847 AGENCOURT
	6	311	12.4	3095	4	CR858326	CR858326 Pongo pyg
c	7	307.2	12.3	5797	4	CR749233	CR749233 Homo sapi
c	8	301.2	12.1	1605	4	CR860521	CR860521 Pongo pyg
	9	301.2	12.1	1863	4	CR599842	CR599842 full-leng
	10	300.6	12.0	946	5	BQ958903	BQ958903 AGENCOURT
c	11	298	11.9	918	5	BQ706343	BQ706343 AGENCOURT
	12	293.6	11.7	736	8	CX780936	CX780936 HESC3_2_B
	13	290.6	11.6	3092	4	CR613629	CR613629 full-leng
c	14	288.2	11.5	855	5	BQ681302	BQ681302 AGENCOURT
c	15	287.6	11.5	769	5	BQ710315	BQ710315 AGENCOURT
c	16	285.2	11.4	4088	4	HSM808459	BX648311 Homo sapi
	17	283	11.3	3165	4	CR859576	CR859576 Pongo pyg
c	18	281.6	11.3	800	1	AU120942	AU120942 AU120942
	19	281.4	11.3	3259	4	CR860168	CR860168 Pongo pyg
	20	280.4	11.2	672	6	CA431692	CA431692 UI-H-DF0-
	21	280	11.2	3474	4	HSM800886	AL110229 Homo sapi
	22	278.8	11.2	658	9	AQ393450	AQ393450 CITBI-E1-
c	23	277.6	11.1	877	9	AQ739838	AQ739838 HS_5505_A
	24	277.2	11.1	1875	4	BC009270	BC009270 Homo sapi
c	25	276.6	11.1	3990	4	HSM803026	AL713681 Homo sapi
	26	276.4	11.1	4828	4	HSM802759	AL390128 Homo sapi
c	27	276.2	11.1	881	5	BU521286	BU521286 AGENCOURT
	28	274.8	11.0	3552	4	CR860263	CR860263 Pongo pyg
	29	273.8	11.0	736	6	CA427039	CA427039 UI-H-DF0-
c	30	273.4	10.9	629	5	BX509360	BX509360 DKFZp686J
c	31	273.4	10.9	2230	4	CR859082	CR859082 Pongo pyg
	32	272.4	10.9	652	6	CA427045	CA427045 UI-H-DF0-
	33	272.4	10.9	666	6	CA431783	CA431783 UI-H-DF0-
c	34	272.4	10.9	815	8	CX785622	CX785622 HESC3_49_
c	35	272.2	10.9	922	5	BU501973	BU501973 AGENCOURT
c	36	272	10.9	617	3	BI861844	BI861844 603388872
	37	271.8	10.9	644	7	CN480313	CN480313 UI-H-EU0-
	38	271.6	10.9	1042	5	BX377759	BX377759 BX377759
	39	271.6	10.9	2097	4	CR614786	CR614786 full-leng
c	40	271.4	10.9	603	1	AL707313	AL707313 DKFZp686P
c	41	271.2	10.9	5325	4	HSM804049	AL832738 Homo sapi
c	42	271.2	10.9	5785	4	HSM802309	AL137734 Homo sapi
c	43	270.6	10.8	2330	4	CR615928	CR615928 full-leng
	44	270.4	10.8	618	9	BZ609884	BZ609884 WHACH76TF
	45	270	10.8	665	5	BU633001	BU633001 UI-H-DF0-

ALIGNMENTS

RESULT 1
BU785040/c
LOCUS BU785040 507 bp mRNA linear EST 11-OCT-2002
DEFINITION in43g09.x1 HR85 islet Homo sapiens cDNA clone IMAGE:6125008 3', mRNA sequence.
ACCESSION BU785040
VERSION BU785040.1 GI:23830576
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 507)
AUTHORS Melton,D., Brown,J., Kenty,G., Permutt,A., Lee,C., Kaestner,K., Lemishka,I., Scearce,M., Brestelli,J., Gradwohl,G., Clifton,S., Hillier,L., Marra,M., Pape,D., Wylie,T., Martin,J., Blistain,A., Schmitt,A., Theising,B., Ritter,E., Ronko,I., Bennett,J., Cardenas,M., Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T., Jackson,Y. and Bowers,Y.
TITLE Endocrine Pancreas Consortium
JOURNAL Unpublished (2000)
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohp.harvard.edu
Library was constructed by Dr. Hiroshi Inoue DNA sequencing by:
Washington University Genome Sequencing Center For information on obtaining a clone please contact: Dr. Hiroshi Inoue
(hinoue@im.wustl.edu)
Seq primer: -40UP from Gibco
High quality sequence stop: 443.
FEATURES Location/Qualifiers
source 1. .507
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6125008"
/tissue_type="Purified pancreatic islet"
/lab_host="DH10B"
/clone_lib="HR85 islet"

```
                     /note="Organ: Pancreas; Vector: pBluescript SK(-); Site_1:
                     NotI; Site_2: XhoI; cDNA made by oligo-dT priming.
                     Size-selected on agarose gel. Average insert size ~1kb. 5'
                     XhoI site was destroyed after directional cloning.
                     Amplified once. Contact information: Hiroshi Inoue, MD,
                     Metabolism Div. (Alan Permutt Lab), Washington University
                     School of Medicine, Box 8127, 660 South Euclid Ave., St.
                     Louis, MO 63110, E-mail: hinoue@imgate.wustl.edu, Tel:
                     314-362-1916, Fax: 314-747-2692."
ORIGIN

  Query Match             20.2%;  Score 505.4;  DB 5;  Length 507;
  Best Local Similarity   99.8%;  Pred. No. 2.3e-35;
  Matches  506;  Conservative    0;  Mismatches    1;  Indels    0;  Gaps    0;

Qy       1506 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 1565
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        507 TACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTT 448

Qy       1566 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 1625
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        447 TAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACAC 388

Qy       1626 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 1685
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        387 ACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCT 328

Qy       1686 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 1745
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        327 GCATTAGAGGATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAA 268

Qy       1746 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 1805
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        267 GCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTC 208

Qy       1806 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 1865
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        207 GAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTG 148

Qy       1866 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCT 1925
             ||||||||||||||||||||||||||||||||||||||||||||||||||| ||||||||||
Db        147 GGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATTGCT 88

Qy       1926 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 1985
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          87 TGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGG 28

Qy       1986 CGACACAGCGAGACTCTATCTCAAAAA 2012
             |||||||||||||||||||||||||||
Db          27 CGACACAGCGAGACTCTATCTCAAAAA 1


RESULT 2
BC024593
LOCUS       BC024593                4087 bp    mRNA    linear   HTC 29-JUN-2004
DEFINITION  Homo sapiens cDNA clone IMAGE:3914314, with apparent retained
            intron.
ACCESSION   BC024593
VERSION     BC024593.1  GI:22137609
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 4087)
  AUTHORS   Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
            Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
            Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
            Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
            Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
            Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
            Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
            Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
            Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
            McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
            Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
            Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
            Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
            Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
            Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
            Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
            Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
            Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
  TITLE     Generation and initial analysis of more than 15,000 full-length
            human and mouse cDNA sequences
  JOURNAL   Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
   PUBMED   12477932
REFERENCE   2  (bases 1 to 4087)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-MAR-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
  REMARK    NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Sequencing Group at the Stanford Human Genome
            Center, Stanford University School of Medicine, Stanford, CA  94305
            Web site:       http://www-shgc.stanford.edu
            Contact:  (Dickson, Mark) mcd@paxil.stanford.edu
            Dickson, M., Schmutz, J., Grimwood, J., Rodriquez, A., and Myers,
            R. M.

            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAK Plate: 22 Row: m Column: 14
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1. .4087
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3914314"
                     /tissue_type="Uterus, leiomyosarcoma"
                     /clone_lib="NIH_MGC_71"
                     /lab_host="DH10B"
                     /note="Vector: pCMV-SPORT6"
ORIGIN

  Query Match             13.8%;  Score 344.2;  DB 4;  Length 4087;
  Best Local Similarity   58.3%;  Pred. No. 6.6e-22;
  Matches  859;  Conservative    1;  Mismatches  544;  Indels   69;  Gaps   12;

Qy        599 TTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 658
              || |   ||| ||||||  ||||| ||||||||||||| | |||| |||||||||  |||
Db       2019 TTTGAAATTTGTTTGTTTGTGTTTTGAGACAGGGTCTGGGTCTGCCACCCAGGCTGGAGT 2078

Qy        659 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 718
              |||||||  | ||||| ||| ||||||| |||||| | | || |||| |||   || | |
Db       2079 GTAGTGGTGTGATCATAGCTCAATGCAGCCCCAACCTTCCGGGCGCAAGTGATGCTCTCA 2138

Qy        719 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 778
              ||||||||||  |||||||||||||||  || ||  ||||||| ||   |||||  | |
Db       2139 CCTCAGCCTCCTGAGTAGCTGGGACTACAGGCATGTGCCACCACACCCTCCTAATATATA 2198

Qy        779 AAAT-----TTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 833
                  |     |||||||||||||   ||| | | || |||| |||| || ||||  |||||
Db       2199 TTTTTTTTATTTTTTGTAGAGATGAGGTTTCATTATGTTGCCCAGACTGGTCTGGAACTC 2258
```





```
RESULT 3
HSM803026
LOCUS       HSM803026              3990 bp    mRNA    linear   HTC 16-APR-2005
DEFINITION  Homo sapiens mRNA; cDNA DKFZp761O0217 (from clone DKFZp761O0217).
ACCESSION   AL713681
VERSION     AL713681.1  GI:19584382
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3990)
  AUTHORS   Ansorge,W., Krieger,S., Regiert,T., Rittmueller,C., Schwager,B.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (22-SEP-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by EMBL (European Molecular Biology Laboratories,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp761O0217) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp761O0217
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3990
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="RZPD:DKFZp761O0217"
                     /db_xref="taxon:9606"
                     /clone="DKFZp761O0217"
                     /tissue_type="amygdala"
                     /clone_lib="761 (synonym: hamy2). Vector pSport1; host
                     DH10B; sites NotI + SalI"
                     /dev_stage="adult"
                     /note="cytochrome c"
     gene            1. .3990
                     /gene="DKFZp761O0217"
     CDS             61. .378
                     /gene="DKFZp761O0217"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAD28485.1"
                     /db_xref="GI:19584383"
                     /db_xref="GOA:P99999"
                     /db_xref="InterPro:IPR002327"
                     /db_xref="InterPro:IPR003088"
                     /db_xref="UniProt/Swiss-Prot:P99999"
                     /translation="MGDVEKGKKIFIMKCSQCHTVEKGGKHKTGPNLHGLFGRKTGQA
                     PGYSYTAANKNKGIIWGEDTLMEYLENPKKYIPGTKMIFVGIKKKEERADLIAYLKKA
                     TNE"
ORIGIN
```

```
  Query Match           13.6%;  Score 341;  DB 4;  Length 3990;
  Best Local Similarity   57.5%;  Pred. No. 1.3e-21;
  Matches  856;  Conservative    1;  Mismatches  541;  Indels   91;  Gaps   10;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
Db       2522 TTTTCTCATTTATTTTTTCTGAGACGGAGTTTCGCCCTTTTCGCCCAGGCTGAGTACAGT 2581

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
Db       2582 GGTGCAATC-----TCACTGCAACCTCCGCCTTCCAGTTGCAAGTGATTCTCCTGCCTCA 2636

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
Db       2637 GCCTCCTGAGTAGCTGGGATTGCAGGCACCCGCCACCACGCCTGGCTAATTTTTGTA--T 2694

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
Db       2695 TTTTAGTGGAGATGGTGTTTCACCATGTTGGCCAGACTGGTCTTGGACTCCTGACCTC-- 2752

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
Db       2753 GTGATCCACCCACCTTGGCCTCCCAAAGTTCTAGGATTACAGGTGTGAGCCACCACGTCC 2812

Qy        904 GGCTACTTATT---------------------------------------------------- 914
Db       2813 GGCCAATTTTTCTCATTTCTATGCCTCCTATATTAAGGTCTGTGTTGGCACAGATGAGTA 2872

Qy        915 ------TCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACT 968
Db       2873 ACTGCCATGTTCTAGGTCAGTTATACCCAAGCACTTCTGGTGGTTTAAAATGTGATTCTG 2932

Qy        969 GCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACC------------ 1016
Db       2933 TAACTTTTTTATTTTTATTTTTTTGAGATAATTTCACTCTTGTTGCCCAGGCTGGAGTGC 2992

Qy       1017 -TCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCC 1075
Db       2993 AATGGCGTGATCGCTGCTCACCGCAACCTCCGCCTTCCAGGTTCAAGCGATTCTCCTGAC 3052

Qy       1076 TAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGT 1135
Db       3053 TCAGCCTCTCAAGTAGCTGGGATTACAAGCATGCGCCACCATGCCCAGCTTATTT--TGT 3110

Qy       1136 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195
Db       3111 GTTTTTAGTAGAGACAGGGTTTCTCCATGCTGGACAGGCTGGTCTTGAACTCCCGACGTC 3170

Qy       1196 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1255
Db       3171 GGGTGATCTGCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCGTGAGCCACCACGC 3230

Qy       1256 CTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCCCCAGATCTAGAGCAGTGTCTGA 1315
Db       3231 CTGGCCAATTA---TGTAATTTTTTAAAAAGGACATTTCTATCAGGGATATATACCTTCA 3287

Qy       1316 CATATAGTAGGTGCTCAATAAATAATTGATGAATGCACAGCCTAGATATAAACTTTCTTT 1375
Db       3288 GAAATAAGGAAATAGGGGAAAAAAAGAGCACTATAAACCACATGTTTTCATTTCTAGTGC 3347

Qy       1376 TTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATAAATACAATCTTGCATTCTGCTT 1435
Db       3348 TTCGCTGTAAGTGGCTAGGTTGGTAGAATCAAAAACAAGGGCCAGATGTATTTAAGGGGT 3407

Qy       1436 TTTCACTTATCACCTTGTTATGACTTTTTCATATTGCCTCAAACCTTTATTGTTACTGTT 1495
Db       3408 ATTCAGATGCCACCTACATGCTTATTTTGTCTAGAACAGTGCTGTCTAATAGAACTTTCT 3467

Qy       1496 TTTTCATTGTTACTATTTTAGTCACTGAATAATATGGCTTAATTTGCTTATACATCCTCC 1555
Db       3468 GTGACGATGGATATTTTGTAGACTTT------TGCTGTCCAGTGTGGTAGCCACTAACCA 3521

Qy       1556 TGCTCCACTTTAGAAGGCCAAATTTACAAATCTGATGAAAGCTATGAACCCTCTCCCCAG 1615
Db       3522 CATGTGGCTGTTAAGCCCTTGAAATATAGCTAGTGTGACTAGAAAGTATTTTATTTTAAA 3581

Qy       1616 AGAAATACACACACACACACACTCACACACAGTTTTTTTTTAATGTTTGCAACTAAGA 1675
Db       3582 TTTACATAGGCACAAGTGGCTAGTGGCTACTGTATTGACATTCTGGGTCTAGGACTAGAA 3641

Qy       1676 CAAGAAACCTGCATTAGAGG--ATGTTTGTTCATATTAATTAAAAATAACTCAGTTGGGC 1733
Db       3642 CCAGTGGTCTGTAACAAAAGTACTTTCTCTTTACTCTATTAAATCTAGAATTAGCCGGGC 3701

Qy       1734 ACAGTGACTCAAGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGA 1793
Db       3702 ATGGTCGCTCATGCCTGTGGTCCCAGCACTTTGGGAGGCCAAGGCAGGCAGATCACTTGA 3761

Qy       1794 GGTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC 1853
Db       3762 GGTCGGGCGTTTGAGACCAGCCTGGTCAACATGGCGAAACCCTGTCTCTACAAAAAACAT 3821

Qy       1854 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1913
Db       3822 AAAAGTTAGCCAGGTGTGGTGGTGGGCACCTGTAGTCTCAGCTACTTGGGAGGCTGAGGC 3881

Qy       1914 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
Db       3882 ACAAGAATCACTTGAACTTGGGAGGTGGAGGTTGCAGTGAGCCAGGATTGTGCCACTGCA 3941

Qy       1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAA 2022
Db       3942 CTCAAGCCTGGGTGAC-GAGTGAAACTGTCTCCAAAAAAAAAAAAAAA 3989
```

```
RESULT 4
AQ839852/c
LOCUS       AQ839852               15970 bp    DNA     linear   GSS 31-MAR-2000
DEFINITION  260L13-C56 CITB Homo sapiens genomic clone 260L13, genomic survey
            sequence.
ACCESSION   AQ839852
VERSION     AQ839852.1  GI:6652484
KEYWORDS    GSS.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 15970)
  AUTHORS   Carpten,J.D., Makalowska,I., Robbins,C.M., Scott,N., Sood,R.,
            Connors,T.D., Bonner,T.I., Smith,J.R., Faruque,M.U., Stephan,D.A.,
            Pinkett,H., Morgenbesser,S.D., Su,K., Graham,C., Gregory,S.G.,
            Williams,H., McDonald,L., Baxevanis,A.D., Klingler,K.W. and
            Landes,G.M.
  TITLE     A 6-Mb high-resolution physical and transcription map encompassing
            the hereditary prostate cancer 1 (HPC1) region
  JOURNAL   Genomics 64 (1), 1-14 (2000)
   PUBMED   10708513
COMMENT     Contact: Carpten JD
            Cancer Genetics Branch
            National Human Genome Research Institute/National Institutes of
            Health
            Bldg. 36, Room 3D04, 36 Convent Drive, Bethesda, MD
            Tel: 301 435 5626
            Fax: 301 435 5465
            Email: jdc@nhgri.nih.gov
            Class: shotgun.
FEATURES             Location/Qualifiers
     source          1. .15970
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /clone="260L13"
                     /clone_lib="CITB"
ORIGIN
```

```
Query Match           13.5%;  Score 337.2;  DB 9;  Length 15970;
Best Local Similarity  57.8%;  Pred. No. 1.1e-21;
Matches  905;  Conservative    1;  Mismatches  529;  Indels  131;  Gaps   12;

Qy       598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657
Db      6787 TTTTTATTTTTATTTATTTTTTTTTTGAGGTGGAGTATCGCTATGTCACCCAGGCTGGAG 6728

Qy       658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717
Db      6727 TGCAGTGGCATGATATCAGCTCACTGCAACCTTCACCTGCTAGATTCAAGCAGTTCTCGT 6668

Qy       718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 777
Db      6667 GACTCAGCCTCCTGAGTAGCTGTGATTACAGGCACATGCCACCATGCTCAGCTAACT-TT 6609

Qy       778 AAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 837
Db      6608 TGTATTTTTTAGTAGAGACAGGGTCTCACCATGTTGGCCAGGCTGGTCTGGAACTCCTGA 6549

Qy       838 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 897
Db      6548 CCTCAAGTGATCCGCCCTCCTTGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACT 6489

Qy       898 ATGTGCGGCTACTTATTTCTTT--------ACATTCCATCTTTCCAATAGAATGTAAGAT 949
Db      6488 GCGCCCAGCCTCTGAGTTCTTTTAAGTCAGAAACTTGAGCTCTTCAGCTCTGATAAATTT 6429

Qy       950 CCACAGAACAGGGATTACTGCCT--ATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007
Db      6428 GGGGGGCAAGGGGACTAATTTTTTCTTTTTCTTTTTCTTTTTTTTAAGATGGAGTCTTGC 6369

Qy      1008 TTCATCACCTCAACCTCCGTTCA--------------GCTCACTGCAACCTCTGCCTCCC 1053
Db      6368 TGTGTTGCCCAGGGTGGAGTGCAGTGGTGTGATCTTGGCTCACTGCAACCTCTGCCTCCC 6309

Qy      1054 GGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCA 1113
Db      6308 AGGTTCAAGCAATTCTCCTGCCTCAGCTTACTGAGTAGCTGGGACTGCAGGCCTGTGCCA 6249

Qy      1114 CCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGG 1173
Db      6248 CCACTCTCAGCTAA-TTTTTGTATTTTTAGTAGAGACAGGGTTTCAGCACATTGGCCAGG 6190

Qy      1174 CTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAA 1233
Db      6189 CTGGTCTCGATCTCCTGACCTCAAGTGATCTGCCTACCTCGTCTCCCCAAAGTGCTGGGA 6130

Qy      1234 TTATAGGCGTGAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTATTGCTATATCC 1293
Db      6129 TTACAGGCAAGAGCCACCACACCTGGCCCTT---GGGGGATGTTATTTCTTTGACAAGTT 6073

Qy      1294 CCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATTGATGAATGCAC 1353
Db      6072 TTGCCTTTCAAATTATATCTGTTGTCTCTTTCAGGAACTCTGTTTAGTTATATTTTGGGT 6013

Qy      1354 AGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAACTTTGCAGAATA 1413
Db      6012 CTTCTAGATTAATCCTTTAATTTTTTAAAATATCTATACGGTTCATCTCTTTGGCAATTA 5953

Qy      1414 AATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTTTTCATATTGCC 1473
Db      5952 GTTCTACTTTATACTTTTTCCTTAATTTTTATTTTCCAACTCTTATTTAAATTTCTGTCA 5893

Qy      1474 TCAAACCTTTATT-----------GTTACTGTTTTTTCATTGTTACTATTTTAGTC---- 1518
Db      5892 TATTTTGTTCATTTCTAAGAGTTATTTCATATTTTTTCACTGTTCCTTTTTTTTTCTTTA 5833

Qy      1519 ---------ACTGAATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGA 1569
Db      5832 GGCTAGTCAAGTGATTATTGTTCCTTTTTTAATAGTGTCATATTGTTTCAGGGATACAAA 5773

Qy      1570 A-------------------------------GGCCAAATTTACAAATCTGATGAAAGC 1597
Db      5772 ATCTCTTACCTTTCTAAGAATTGATTATACTTGGGTATATTTTATATTTAGGGGAGAATA 5713

Qy      1598 TATGAACCCTCTCCCCAGAGAAATACACACACACACACACTCACACACAGTTTTTTTT 1657
Db      5712 GGGGTTTTTTGTTTGTTTTGTATTGTTGCTGTTGTTTTGTTGTTCAGTTCTCCTGTTTTT 5653

Qy      1658 TAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTGTTCATATTAATTAAA 1717
Db      5652 TAACTTTTTCTGTTTTGTAATTTTGTTCTCTGTCTGTCATGGTAGTTCTCAAATGTTTTG 5593

Qy      1718 AATAACTCAGTTGGGCACAGTGACTCA---------------------------------- 1744
Db      5592 TGGTCTTCGGTTGTCCACAGTGAAAAATTGTTTTAAAGCACTTGGGGCAGAGCTTATAAA 5533

Qy      1745 ----------AGCCTGTAACCACAGTACTTTGGAAGTCCAAGGTGGGTGGATCACTTGAG 1794
Db      5532 CCAACGGATTTGGCTGTAATCCCAGCACTTTGGGAGGCTGAGGCGGGCAGATCA--TGAG 5475

Qy      1795 GTGAGAAGTTCGAGACCAGCCTGGTCAATATGGTGAAACCCTATCTCTACTAAAAATAC- 1853
Db      5474 GTCAGGAGATCGAGACCATCCTGGCTAACACAGTGAAACCCTGTCTCTACTAAAAATACA 5415

Qy      1854 AAAAATTAGCTGGGTGTAGTGATGCATGCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGC 1913
Db      5414 AAAAATTAGCCGGGCGTGGTGGCGGGCACCAATAGTCCCAGCTACTCGGGAGGCTGAGGC 5355

Qy      1914 AAGAGAATTGCTTGAACCTGGGAGGCAGAGGTTGCAGTGAGCCGAGATCCCACCACTGCA 1973
Db      5354 AGGAGAATGGCGTGAACCTGGGAGGCGGAGCTTGCAGTGAGCCAAGATCGCACCACTGCA 5295

Qy      1974 CTCCAGCCTGGGCGACACAGCGAGACTCTATCTCAAAAAAATAAATAAATAAAATAAAGG 2033
Db      5294 CTTGAGCCTAGGCAACAAAGCGAGACTGTCTCAAAAAAAAAATAAAAATAAAAATAAACC 5235

Qy      2034 ATCGGA 2039
Db      5234 AATGGA 5229
```

```
RESULT 5
CD557847/c
LOCUS       CD557847                897 bp    mRNA    linear   EST 11-JUN-2003
DEFINITION  AGENCOURT_14423258 NIH_MGC_180 Homo sapiens cDNA clone
            IMAGE:30390116 5', mRNA sequence.
ACCESSION   CD557847
VERSION     CD557847.1  GI:31583915
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 897)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Michael Brownstein
             cDNA Library Preparation: Invitrogen Corp
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: NDAM464  row: k  column: 21
            High quality sequence stop: 636.
FEATURES             Location/Qualifiers
     source          1. .897
```

```
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30390116"
                     /lab_host="DH10B-Ton A ( T1 and T5 phage resistances)"
                     /clone_lib="NIH_MGC_180"
                     /note="Organ: Testis; Vector: pCMV-SPORT6.1; Site_1: NotI;
                     Site_2: EcoRV (destroyed); Library is oligo-dT primed and
                     directionally cloned (EcoRV site is destroyed upon
                     cloning). Average insert size 1.68 kb. Library was
                     constructed by (Invitrogen). Note: this is a NIH_MGC
                     Library."
ORIGIN

  Query Match             12.6%;  Score 315.8;  DB 6;  Length 897;
  Best Local Similarity   68.1%;  Pred. No. 5.1e-19;
  Matches  469;  Conservative    1;  Mismatches  213;  Indels    6;  Gaps    2;

Qy        583 TGTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTG 642
              |||| |       |  ||   || |||| | ||  | ||| ||||| |||||  |||||
Db        765 TGTATGCATCCCGATATTATTTTATTTTATTTTTATTTTTTGAGACATGGTCTCACTCTG 706

Qy        643 TCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGC 702
              |  ||||| |  |||  ||||||    | |  || |||||||| |||| ||||||   |
Db        705 TTGCCCAGCCTGGAGTGCAGTGGCAGGATCTCAGCTCACTGCAACCTCCACCTCCCAGGT 646

Qy        703 TCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCAC 762
              |||||  ||   | |||||||||||||||||||||||||| |||  | ||||||||||||
Db        645 TCAAGCAATTCTCCCACCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCAC 586

Qy        763 GCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTT 822
              |||  |||||||         ||||  ||| |||  |||| || | | | ||||||||||
Db        585 GCCCAGCTAATTTTTTTGTATTTTAAGTACAGATGGGGTTTTGCCATGCTGGCCAGGCTC 526

Qy        823 GTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTA 882
              ||| ||| ||||| | |||||  |||  || |||||||| ||||||||||||||||||||
Db        525 CTCTCAAATTCCTGACCTCAAGTGATCTGCCCACCTTGGCCTCCCAAAGTGCTGGGATTA 466

Qy        883 CAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAAT 942
              |||| |||||||||| ||  | ||| | |  |  | |  | | | |       | | || |
Db        465 CAGGCGTGAGCCACCGTGCCCAGCTGCATCCTGATATTTAATGCTTGCCAGAATTTGATT 406

Qy        943 GTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGT 1002
              | |  || |  |           | |   | |||| ||     |       | | |     |
Db        405 GCACCATGCTTATTGGCTTCTATCTTTTTTTTTTTTTTTGAGATGGAGACTCACTCTCTCT 346

Qy       1003 CTCACTTCATCACCTCAACCTCCGT----TCAGCTCACTGCAACCTCTGCCTCCCGGGTT 1058
              ||| |    |   | |      ||    |||||||||||||||||  | ||| |||||
Db        345 CTCCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAACCTCCACTTCCTGGGTT 286

Qy       1059 CAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATG 1118
              ||||:|||||||||| || ||||||| |||||||||| | ||||   |||  ||||||
Db        285 CAAGCGATTCTCCTGTCTCAGCCTCCAGAGTAGCTGGGACTACAGATGTGTGCCACCACA 226

Qy       1119 CTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGT 1178
              | |||||||  ||| |||||||||| |||||||||||| |||||||| |||| ||  |
Db        225 CCTGGCTAA--TTTCGTATTTTTAGTAGAGATGGGGTTTCACCATGTTGGCCAGACTAAT 168

Qy       1179 CTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATA 1238
              ||| |||||||||||||| ||||||  ||||||||   | ||||||||||| |||  ||| |
Db        167 CTCGAACTCCTGACCTCAGGTGATCCGCCTGCCTTGGCCTCCCAAAGTGTTGGGGTTACA 108

Qy       1239 GGCGTGAGTCACTGTGCCTGGCCGATTAC 1267
              |||||||| |||    || | | | || |
Db        107 GGCGTGAGCCACCACCCCAGCCGGCTTCC 79


RESULT 6
CR858326
LOCUS       CR858326               3095 bp    mRNA    linear   HTC 12-NOV-2004
DEFINITION  Pongo pygmaeus mRNA;. cDNA DKFZp469P012 (from clone DKFZp469P012).
ACCESSION   CR858326
VERSION     CR858326.1  GI:55727613
KEYWORDS    HTC.
SOURCE      Pongo pygmaeus (orangutan)
  ORGANISM  Pongo pygmaeus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Pongo.
REFERENCE   1  (bases 1 to 3095)
  AUTHORS   Bloecker,H., Boecher,M., Brandt,P., Mewes,H.W., Weil,B., Amid,C.,
            Osanger,A., Fobo,G., Han,M. and Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by GBF (National Research Centre for Biotechnology Ltd.,
            Braunschweig/Germany) within the cDNA sequencing consortium of the
            German Genome Project.
            This clone (DKFZp469P012) is available at the RZPD Deutsches
            Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
            Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp469P012
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .3095
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp469P012"
                     /tissue_type="kidney"
                     /clone_lib="469 (synonym: pkid1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="hypothetical protein (Homo sapiens)"
     gene            1. .3095
                     /gene="DKFZp469P012"
     CDS             120. .533
                     /gene="DKFZp469P012"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH90562.1"
                     /db_xref="GI:55727614"
                     /translation="MLSGRLVPGLVSMAGRVCLCQGSAGSGAVGPVEAAIRTKLEQAL
                     SPEVLELRNESGGHAVPPGSETHFRVAVVSSRFEGLSPLQRHRLIHAALAEELAGPVH
                     ALAIQARTPAQWGENSQLDTSPPCLGGNKKTLGTP"
ORIGIN

  Query Match             12.4%;  Score 311;  DB 4;  Length 3095;
  Best Local Similarity   56.6%;  Pred. No. 6.1e-19;
  Matches  883;  Conservative    1;  Mismatches  511;  Indels  166;  Gaps   10;

Qy         584 GTAAGTGGTGACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGT 643
               || ||      ||   | | ||||||  |||  | |  |||  ||||  | ||||  || |||
Db        1570 GTGAGCCACTACCTGTAGCGTTTTTGTTTTGTTTGTTTTTGAGATGGAGTCTCACTTTGT 1629

Qy         644 CACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCT 703
               ||||  |  |||  |||||||| | |  || ||||||||||||||  |||||    | |
Db        1630 TGCCCAACCTGGAGTGCAGTGGTGCGATCTCAGCTCACTGCAGCCTCCGCCTCCCAGGTT 1689

Qy         704 CAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACG 763
               ||||  ||   |  ||||||||||||||||||||||||||||||||  || ||  |||||||
Db        1690 CAAGCAATTCTCCTGCCTCAGCCTCCCAAGTAGCTGGGACTAGAGGCTTGTGCCACCACA 1749

Qy         764 CCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTG 823
               ||  |||||||         | |||| | ||||| |||| | || | |||| |||||||  |
Db        1750 CCCAGCTAATT---TTTTGTATTTTATGGAGACAGGGTTTCACCATGTTGCCCAGGCTGG 1806

Qy         824 TCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTAC 883
               ||||| |||||||   ||| ||| ||| || ||||  || ||||||  ||||| |||||||
Db        1807 TCTTAGACTCCTGAGCTCAGGCACTCCGCCCACCTCAGCCTCCCAAGGTGCTAGGATTAC 1866

Qy         884 AGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACATTC---------------------- 925
```

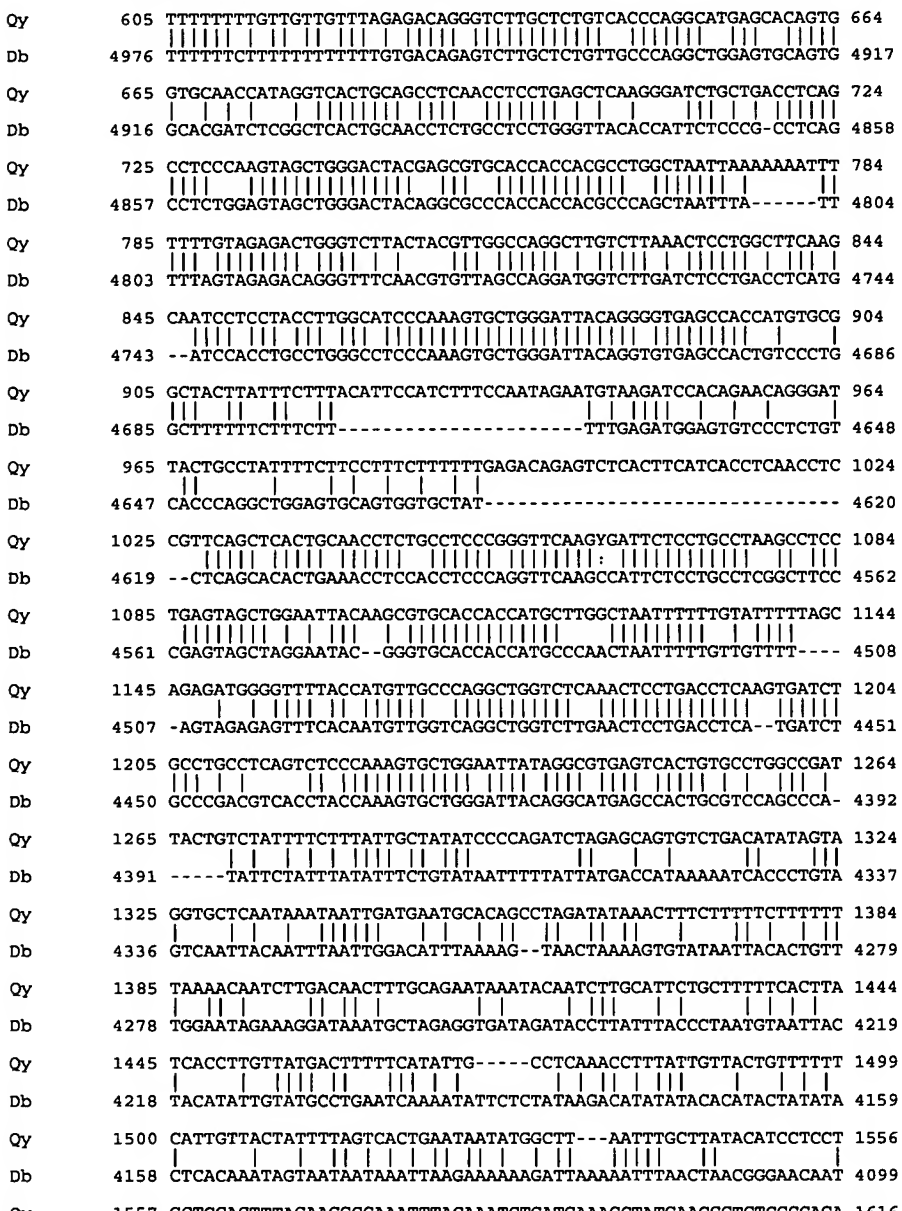
```
Db       1867 AGGCGTAAGCCACCATGCCCTGCAATGTAGCCCTTTTTGAGCATTACAGATATTGCCATA 1926
Qy        926 ----------------------------------CATCTTTCCAAT------------A 938
Db       1927 TTTCTGAGGGGAGCTCAAGTAATGTCTATAGAAGCAACTTTCCAATAATCTCAATGTGGA 1986
Qy        939 GAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACA 998
Db       1987 AATTATTGGAGAACAAAACTTGGTTAACATACCTTTTTTTTTTTTTTTTTTTTGAGACG 2046
Qy        999 GAGT---------------CTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAACC 1043
Db       2047 GAGTTTTGCTCTTGTTGCCCAGACTGGAGTGCAATGGCCCAATCTCAGCTCACTGCAACC 2106
Qy       1044 TCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAA 1103
Db       2107 TCCGCCTCCCAGGTTCAAGCAATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACAG 2166
Qy       1104 GCGTGCACCACCATGCTTGGCTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCAT 1163
Db       2167 GCTCCTGCCACCATGCCTGGCTAA-TTTTTGTATTTTTAGTGGAGACAAGGTTTCATCAT 2225
Qy       1164 GTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAA 1223
Db       2226 ATTGGTCAGGCTGGTCTCAAACTCCTGACCTCAAGCGATCCGCCTGCGTCAACCTTCCAA 2285
Qy       1224 AGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC-GATTACTGTCTATTTTCTTTA 1282
Db       2286 GTTGCTGGGATTACATGCGTGAGCCACCACACCCGGCCTGCTTAACATATATCTTGACCA 2345
Qy       1283 TTGCTATATCCCCAGATCTAGAGCAGTGTCTGACATATAGTAGGTGCTCAATAAATAATT 1342
Db       2346 TTTCTAGCTGAAGAATTTAAGGTGCATGTGGTTGGGCAGTAGGGAGTCATTGACAAATTT 2405
Qy       1343 GATGAATGCACAGCCTAGATATAAACTTTCTTTTTCTTTTTTTAAAACAATCTTGACAAC 1402
Db       2406 TGAGTACAGACATGCCAGAGAGCA-----------TTTGGGGGCAGATGGATCTGGCAGT 2454
Qy       1403 TTTGCAGAATAAATACAATCTTGCATTCTGCTTTTTCACTTATCACCTTGTTATGACTTT 1462
Db       2455 GATGTGTGAGTGAATTGGCATGGGAGGTTGGAGGAACAGAGATCAATCCAGAGTGAAGGT 2514
Qy       1463 TTCATATTGCCTCAAACCTTTATTGTTACTGTTTTTTCATTGTTACTATTTTAGTCACTG 1522
Db       2515 AGAACCCTGACTTACAAGTCATCTGTGGGAATGGAAAAGAAGGAGTTTATTTGAGAGCTA 2574
Qy       1523 AATAATATGGCTTAATTTGCTTATACATCCTCCTGCTCCACTTTAGAAGGCCAAATTTAC 1582
Db       2575 CCATAAAGTCAGAAAATGGTTTGATTGGAACAGAAATGGACTAGAGAGGGGAGTCAATGG 2634
Qy       1583 AAATCTGATGAAAGCTATGAACCCTCTCCCCAGAGAAATACACACACACACACACACTCA 1642
Db       2635 TA-------------------------------------------------------ACCCC 2641
Qy       1643 CACACAGTTTTTTTTTAATGTTTGCAACTAAGACAAGAAACCTGCATTAGAGGATGTTTG 1702
Db       2642 AAGGCAGTGGTTTTCAGATTTTCTCAGGTAAGAGCTCCCATAGGAGATTG---------- 2691
Qy       1703 TTCATATTAATTAAAAATAACTCAGTTGGGCACAGTGACTCAAGCCTGTAACCACAGTAC 1762
Db       2692 ------TTTAAAATGAAGCTTCTGGCTGGGTGTGGTGGCTCACACCTGTAATCCCAGCAG 2745
Qy       1763 TTTGGAAGTCCAAGGTGGGTGGATCACTTGAGGTGAGAAGTTCGAGACCAGCCTGGTCAA 1822
Db       2746 TTTGGGAGGCCATGGCAGGCGGATTGCGTGAGGTCAGGAGTTCGAGACCAGCTTGGGCAG 2805
Qy       1823 TATGGTGAAACCCTATCTCTACTAAAAATACAAAAATTAGCTGGGTGTAGTGATGCATGC 1882
Db       2806 CATGGTGAAACCGTATCTC-ACTAAAAATACAAAAATTAGTTGGGTGTGGTGGCAGCCAC 2864
Qy       1883 CTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAAGAGAATTGCTTGAACCTGGGAGGCAGA 1942
Db       2865 CTGTAATCCCAGCCTCTTAGGCAGCTGAGGCAGGAGAATCACTTGAACCTGGGAGGCAGA 2924
Qy       1943 GGTTGCAGTGAGCCGAGATCCCACCACTGCACTCCAGCCTGGGCGACACAGCGAGACTCT 2002
Db       2925 GGTTGCAGTGAGCCAAGATCACGCCACTGCACTCCAGCCTGGGCAACAAAGTGAAACTCT 2984
Qy       2003 ATCTCAAAAAAATAAATAAATAAAATAAAGGATCGGAGAGAAACAAAACTAATAAGATTC 2062
Db       2985 GTCTCAAAAAAAAAAAAAAAAAAAAAAAAAAACCACGTAGATTCAAGGTCCCTATCCTTC 3044
Qy       2063 C 2063
Db       3045 C 3045
```

```
RESULT 7
CR749233/c
LOCUS       CR749233               5797 bp    mRNA    linear   HTC 19-AUG-2004
DEFINITION  Homo sapiens mRNA; cDNA DKFZp686C0331 (from clone DKFZp686C0331).
ACCESSION   CR749233
VERSION     CR749233.1  GI:51476197
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 5797)
  AUTHORS   Poustka,A., Albert,R., Moosmayer,P., Schupp,I., Wellenreuther,R.,
            Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and
            Wiemann,S.
  CONSRTM   The German cDNA Consortium
  TITLE     Direct Submission
  JOURNAL   Submitted (17-AUG-2004) MIPS, Ingolstaedter Landstr.1, D-85764
            Neuherberg, GERMANY
COMMENT     Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
            sequenced by DKFZ (German Cancer Research Center,
            Heidelberg/Germany) within the cDNA sequencing consortium of the
            German Genome Project. This clone (DKFZp686C0331) is available at
            the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in
            Berlin, Germany. Please contact RZPD for ordering:
            http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp686C0331
            Further information about the clone and the sequencing project is
            available at http://mips.gsf.de/projects/cdna/.
FEATURES             Location/Qualifiers
     source          1. .5797
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="DKFZp686C0331"
                     /tissue_type="testis"
                     /clone_lib="686 (synonym: hlcc3). Vector pSport1_Sfi; host
                     DH10B; sites SfiIA + SfiIB"
                     /dev_stage="adult"
                     /note="hypothetical protein"
     gene            1. .5797
                     /gene="DKFZp686C0331"
     CDS             103. .1725
                     /gene="DKFZp686C0331"
                     /codon_start=1
                     /product="hypothetical protein"
                     /protein_id="CAH18089.1"
                     /db_xref="GI:51476198"
                     /translation="MGPLQFRDVAIEFSLEEWHCLDTAQRNLYRNVMLENYSNLVFLG
                     ITVSKPDLITCLEQGRKPLTMKRNEMIAKPSVMCSHFAQDLWPEQSMKDSFQKVVLRR
                     YEKCEHDNLQLKKGCISVDECKVHKEGYNELNQCLTTTPRKICQCDKYVKVLHQFPNS
                     NGQKRGHTGKKPFKYIECGKAFKQFSTLTTHKKIHTGGKPYKCEECGKAFNHSCSLTR
                     HKKIHTGEKPYKCEECGKAFKHSSTLTTHKRNHTGEKPYKCDKCGKAFMSSSTLSKHE
                     IIHTEKKPYKCEECGKAFNRSSTLTTHKIIHTGEKPYKCEECDKAFKYSYTLTTHKRI
                     HTEDKPYKCEECGKAFKYSSTLTTHKRIHTGEKPYKCEECGKAFKRSSDLTTHKIIHT
                     GEKPYKCEECGKAFKYSSNLTTHKKIHTGERPYKCEECGKAFNQSSILTTHKRIHTGE
                     KFYKCEECGKAFKRSSHLTTHKIIHTGEKFYKCEECDKAFKHFSNLTTHKKIHTGKKP
```

YRCEECGKAFKHSSKLTTHKRFHTGEKPYRCE"

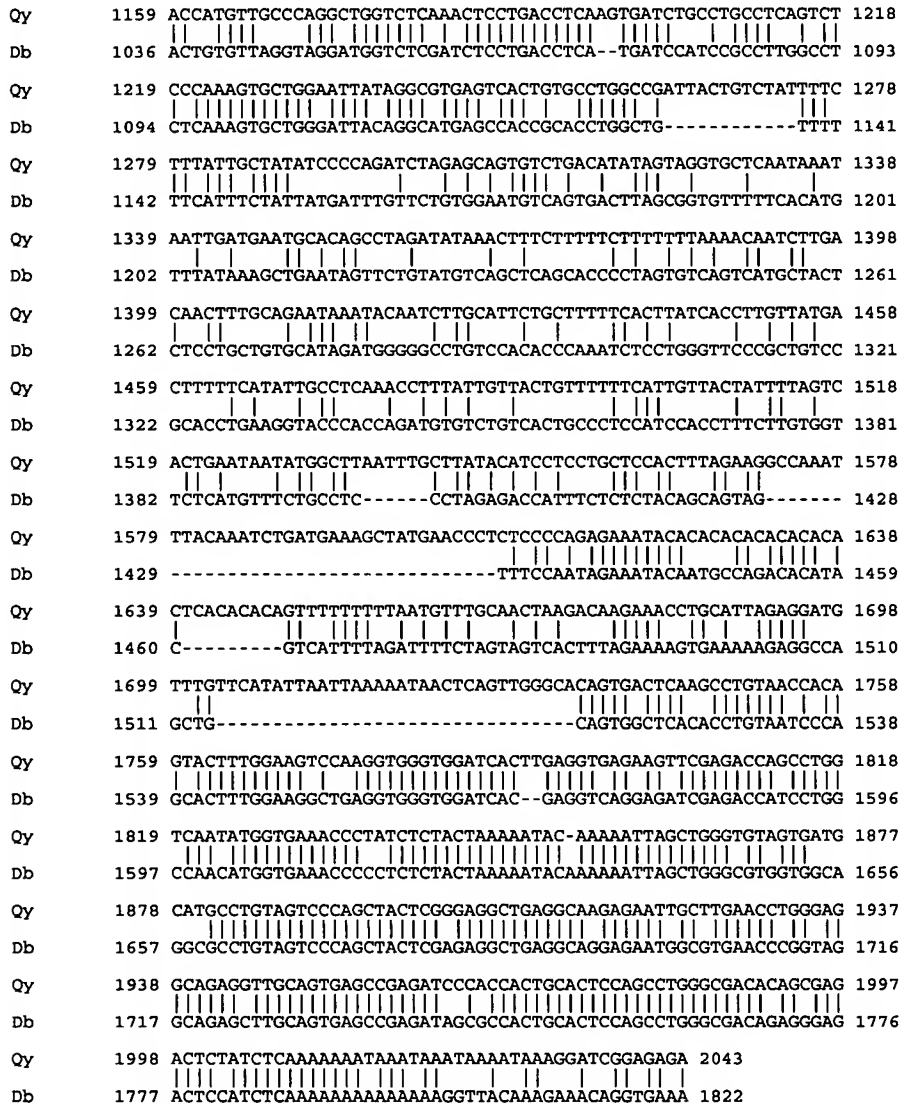
ORIGIN

Query Match 12.3%; Score 307.2; DB 4; Length 5797;
Best Local Similarity 56.9%; Pred. No. 8.8e-19;
Matches 902; Conservative 1; Mismatches 569; Indels 114; Gaps 14;



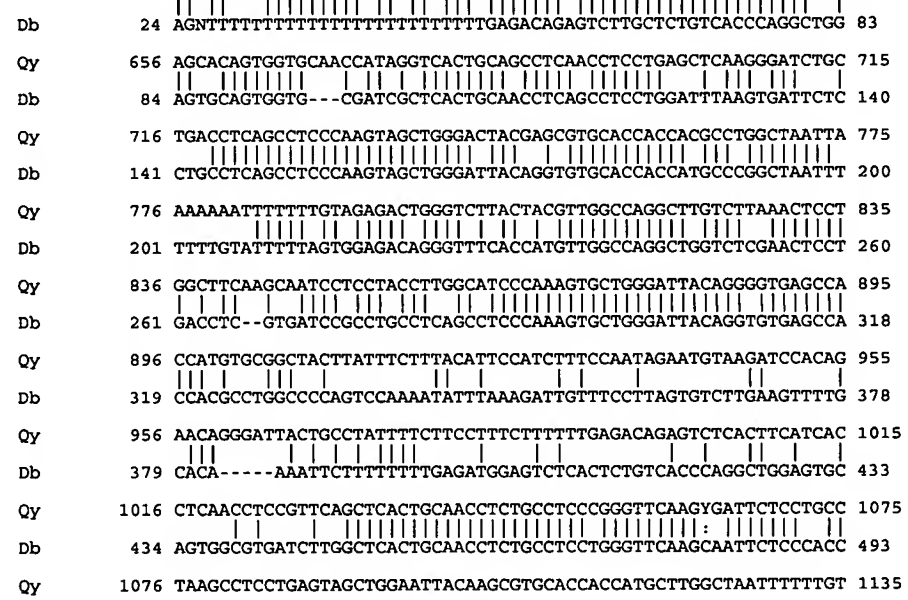
RESULT 8
CR860521/c
LOCUS CR860521 1605 bp mRNA linear HTC 12-NOV-2004
DEFINITION Pongo pygmaeus mRNA; cDNA DKFZp468P2419 (from clone DKFZp468P2419).
ACCESSION CR860521
VERSION CR860521.1 GI:55731876
KEYWORDS HTC.
SOURCE Pongo pygmaeus (orangutan)
ORGANISM Pongo pygmaeus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Pongo.
REFERENCE 1 (bases 1 to 1605)
AUTHORS Ottenwaelder,B., Obermaier,B., Deutschenbaur,S., Schaipp,A., Mewes,H.W., Weil,B., Amid,C., Osanger,A., Fobo,G., Han,M. and Wiemann,S.
CONSRTM The German cDNA Consortium
TITLE Direct Submission
JOURNAL Submitted (12-NOV-2004) MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY
COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by Medigenomix (Martinsried/Germany) within the cDNA sequencing consortium of the German Genome Project.
This clone (DKFZp468P2419) is available at the RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH in Berlin, Germany.
Please contact RZPD for ordering:
http://www.rzpd.de/cgi-bin/products/cl.cgi?CloneID=DKFZp468P2419
Further information about the clone and the sequencing project is available at http://mips.gsf.de/projects/cdna/.

```
FEATURES             Location/Qualifiers
     source          1. .1605
                     /organism="Pongo pygmaeus"
                     /mol_type="mRNA"
                     /db_xref="taxon:9600"
                     /clone="DKFZp468P2419"
                     /tissue_type="heart"
                     /clone_lib="468 (synonym: phrt1). Vector pSport1_Sfi; host
                     DH10B; sites SfilA + SfilB"
                     /dev_stage="adult"
                     /note="unclassified"
ORIGIN

  Query Match             12.1%;  Score 301.2;  DB 4;  Length 1605;
  Best Local Similarity   67.7%;  Pred. No. 6.6e-18;
  Matches  474;  Conservative    1;  Mismatches  204;  Indels   21;  Gaps    3;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657
Db       1602 TTTTTTTTTTTTTGTTTGTTTGTTTTGAGACAGAATCTTGCCTTGTCTCCCGGGCTGGAG 1543

Qy        658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717
Db       1542 TGTAGTGGTGCAATCTCAGCTCACTGCAACATCCATCTCCCGGATTCCACTTATTCTCCT 1483

Qy        718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAA 777
Db       1482 GCCTCAGCCTCTCAGGTAGCTGGGATTATGGGTGCCTGCCATAATGCCCAACTCAATTTT 1423

Qy        778 AAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGG 837
Db       1422 GTA--CTTTTAGTAGAGACGGGGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTCCTGG 1365

Qy        838 CTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACC 897
Db       1364 CCTCAAGTGATCCACCTGCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGACCACTG 1305

Qy        898 ATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAA 957
Db       1304 CACCTGGCCTGAATTTCTCCATTCTTCCCA-----CACACCCGCCTCAGGTTCTCCTTCC 1250

Qy        958 CAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCT 1017
Db       1249 TGATCATTGACGCTTTTTTTTTTTTTTTTTTTTTTGAGACAGCATCTCACTGTCTCACCC 1190

Qy       1018 CAACCTCCGTTCA--------------GCTCACTGCAACCTCTGCCTCCCGGGTTCAAGY 1063
Db       1189 AGACTGGAGTGCAGTAGTGCAATCTTGGCTCACTGAAACCTCTTCCTCCCAGCCTCAAGC 1130

Qy       1064 GATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGG 1123
Db       1129 GATTCTCCTGTCTCAGCCTCCCGAGTAGCTGGGATTACATGTGTGCACCACTAACACCCA 1070

Qy       1124 CTAATTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAA 1183
Db       1069 GATAATTTTTGTACTTTTAGTAGAGATGGGGTTTGACCATATTGGCCAGACTGGTCTTGA 1010

Qy       1184 ACTCCTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGT 1243
Db       1009 ATTCCTGACCTCAGGTGATCCGCCCACCTCAGCCTCCCAAAGTGTTGGGGTTATAGGCGT 950

Qy       1244 GAGTCACTGTGCCTGGCCGATTACTGTCTATTTTCTTTAT 1283
Db        949 GAGCCACTGCACCCGGCCCCCTTCCTTCATCTTAGTCAAT 910


RESULT 9
CR599842
LOCUS       CR599842               1863 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DA002YO06 of Neuroblastoma of Homo
            sapiens (human).
ACCESSION   CR599842
VERSION     CR599842.1  GI:50480649
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 1863)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 1863)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .1863
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DA002YO06"
                     /tissue_type="Neuroblastoma"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             12.1%;  Score 301.2;  DB 4;  Length 1863;
  Best Local Similarity   56.9%;  Pred. No. 6e-18;
  Matches  811;  Conservative    1;  Mismatches  509;  Indels  105;  Gaps   10;

Qy        624 GAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGGTGCAACCATAGGTCACTG 683
Db        496 GGGTCAGAGTCTCACTCTGTTTCCCAGGCTGGAGTTCTGTGGCACCGTGTTGGCTCACTG 555

Qy        684 CAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGCCTCCCAAGTAGCTGGGAC 743
Db        556 CAACCTCCGTCTCCTGGGTTCAAGCAATTCTCTTGCCTCAGCCTCCAGAGTAGCTGGGAT 615

Qy        744 TACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTTTTTGTAGAGACTGGGTCT 803
Db        616 TACAGGTGCCCACCACCACACCCGGCTAATTTTTTTATATTTTTAGTAGACACGGGGTTT 675

Qy        804 TACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGCAATCCTCCTACCTTGGCA 863
Db        676 CACCATGTTGGCCAGGCCGGTCTTGAACTTGTGACGTCAGGTGATGCACCCGCCTTGGCC 735

Qy        864 TCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGGCTACTTATTTCTTTACAT 923
Db        736 TCCCAAAGTGCTGGGATTACAGGTCCAGATTTTGAACAAATCTCGCTAGAGATTGAACAG 795

Qy        924 TCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATTACTGCCTATTTTCTTCCT 983
Db        796 AGCTGCGCCTGGATCCCGCCACAGTGAGGAGACCTGAAGACCAGAGAAAACACAGCAAAC 855

Qy        984 TTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCCGTTCAGCTCACTGCAA-C 1042
Db        856 GGAGTCTTGCTCTGTTGTCCAGGCTGGAGTGTAGTGTGGCGATCTCGGCTCCCTGCAAGA 915

Qy       1043 CTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACA 1102
Db        916 ATCTGCCTCCCAGATTCACGCCATTTTCCCACCTCAGCCCCTGGAGTAGCTGGAAGTACA 975

Qy       1103 AGCGTGCACCACCATGCTTGGCTAAT----TTTTTGTATTTTTAGCAGAGATGGGGTTTT 1158
Db        976 GGCTCCCGCCACCACGCCTGGCTAATTTTGTTTTTGTATTTTTAGTAGAGACGGGGTTTC 1035
```



RESULT 10
BQ958903
LOCUS BQ958903 946 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT_10035485 NIH_MGC_40 Homo sapiens cDNA clone IMAGE:6483214 5', mRNA sequence.
ACCESSION BQ958903
VERSION BQ958903.1 GI:22374381
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 946)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: DCTD/DTP
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2667 row: g column: 23
High quality sequence stop: 534.
FEATURES Location/Qualifiers
source 1. .946
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6483214"
/tissue_type="carcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_lib="NIH_MGC_40"
/note="Organ: prostate; Vector: pOTB7; Site_1: XhoI; Site_2: EcoRI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCACGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH_MGC Library."
ORIGIN

Query Match 12.0%; Score 300.6; DB 5; Length 946;
Best Local Similarity 69.1%; Pred. No. 1e-17;
Matches 468; Conservative 1; Mismatches 197; Indels 11; Gaps 4;



```
Db        494 TCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTTGC 552

Qy       1136 ATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTC 1195

Db        553 ATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGGTGGTCTTGAACTCCTGACCTC 612

Qy       1196 AAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGC 1255

Db        613 AGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGTGC 672

Qy       1256 CTGGCCGATTACTGTCT 1272

Db        673 TCAGCCGCANAATTCCT 689


RESULT 11
BQ706343/c
LOCUS       BQ706343                 918 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8484983 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6301091
            5', mRNA sequence.
ACCESSION   BQ706343
VERSION     BQ706343.1  GI:21845242
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 918)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2516  row: g  column: 12
            High quality sequence stop: 632.
FEATURES             Location/Qualifiers
     source          1. .918
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6301091"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match             11.9%;  Score 298;  DB 5;  Length 918;
  Best Local Similarity   68.6%;  Pred. No. 1.8e-17;
  Matches  456;  Conservative    1;  Mismatches  197;  Indels   11;  Gaps    3;

Qy        598 TTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAG 657

Db        666 TTTTGGATTATGATTATTATTCTTTGAAGCCAGGGTCTCGCTCTGTCACNCAGGCCGGAG 607

Qy        658 CACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTG 717

Db        606 TGCAGTGGTGCAATCATGGCTCACTCCAGCCTTTATCTCCTGGGCCCAAGTGATCCTCCA 547

Qy        718 ACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTA------ 771

Db        546 ACCTCAGCCTCCCAAGTAGCTGGGATTATAAGTGCGTGCTACGATGTCTGGCTAATTTTT 487

Qy        772 -ATTAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAA 830

Db        486 TATTTATAAATATTTTTTGTAGAGACAGGGTCTCGCTTTGTTGCCCAGGCTGGTCTCAAA 427

Qy        831 CTCCTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTG 890

Db        426 TTCCTGGACTCAGGCGATCCTCCTGCTTCAGCCTCCCAAAGTGCTGAGATTACAGGCGTG 367

Qy        891 AGCC---ACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAG 947

Db        366 AACCTTTCAACTTTTAATGCTTTCAGTATTTTCCCATCCCTTGATTGAACATTTGGGTTT 307

Qy        948 ATCCACAGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCAC 1007

Db        306 TAAATTATTTTTGTTGTTGTTGTTTCTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGC 247

Qy       1008 TTCATCACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATT 1067

Db        246 TAGAATGCAGGGGTGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGCGTTCAAGCGATT 187

Qy       1068 CTCCTGCCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA 1127

Db        186 CTTGTGCCTCAGCCTCCTGAGTAGCTGGGATTACAGGCGCCTGCCATGATGCCCAACTAA 127

Qy       1128 TTTTTTGTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTC 1187

Db        126 TTTTTTGTATTTTTAGTAGAGACAGTGTTTCACCATGTTGGCCAGGCTGGTCTCAAACTC 67

Qy       1188 CTGACCTCAAGTGATCTGCCTGCCTCAGTCTCCCA-AAGTGCTGGAATTATAGGCGTGAG 1246

Db         66 CTGACCTCAAGTGATCCGCGCACCTTGGCCTCCCANAATTGCTGGGATTACAGGTGTGAG 7

Qy       1247 TCACT 1251

Db           6 CCACT 2


RESULT 12
CX780936
LOCUS       CX780936                 736 bp    mRNA    linear   EST 02-MAR-2005
DEFINITION  HESC3_2_B04.b1_A036 NIH_MGC_260 Homo sapiens cDNA clone
            IMAGE:30928759 3', mRNA sequence.
ACCESSION   CX780936
VERSION     CX780936.1  GI:58297726
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 736)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Other_ESTs: HESC3_2_B04.g1_A036
            Contact: Daniela S. Gerhard, Ph.D.
            Office of Cancer Genomics
            National Cancer Institute / NIH
            Bldg. 31 Rm10A07 Bethesda, MD 20892
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: BresaGen, Inc.
             cDNA Library Preparation: Express Genomics, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Laboratory for Genomics and Bioinformatics,
            University of Georgia
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
```

```
           Plate: NDAM1172  row: c  column: 08
           Seq primer: M13-21 (TGTAAAACGACGGCCAGT)
           High quality sequence stop: 711
           POLYA=Yes.
FEATURES             Location/Qualifiers
     source          1. .736
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:30928759"
                     /sex="male"
                     /tissue_type="human embryonic stem cells"
                     /cell_type="human embryonic stem cells"
                     /cell_line="BG01"
                     /lab_host="DH10B-T1 phage-resistant E. coli"
                     /clone_lib="NIH_MGC_260"
                     /note="Vector: pExpress-1; Site_1: NotI; Site_2: EcoRV;
                     RNA obtained from human embryonic stem cells isolated from
                     the inner cell mass of blastocyst stage embryos. Cell line
                     id and NIH Registry designation is BG01. Positive for
                     SSEA3, SSEA4, Tra 1-60, Tra 1-81, CD9, Alk Phos, Oct4 and
                     Nanog expression; negative for SSEA1 expression. Passage
                     number 21. cDNA primed using oligo-dT primer:
                     5'-pGACTAGTTCTAGATCGCGAGCGGCCGCCC(T)25-3' and cloned into
                     the EcoRV/NotI sites of pExpress-1. This primary library
                     is non-normalized (normalized primary library is
                     NIH_MGC_261). It was constructed by Express Genomics
                     (Frederick, MD). Sequence ends have been trimmed to
                     exclude vector and regions below Phred quality 16. Note:
                     this is a Mammalian Gene Collection library."
ORIGIN

  Query Match             11.7%;  Score 293.6;  DB 8;  Length 736;
  Best Local Similarity   66.9%;  Pred. No. 5e-17;
  Matches  464;  Conservative    1;  Mismatches  220;  Indels    9;  Gaps    3;

Qy        599 TTGGGTTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGC 658
              ||   |||||||||  || ||  ||| ||||  | ||||  |||| |  |||||||   ||
Db          1 TTTTTTTTTTTTTTTTTTTTTTTTGAGATGGAGTCTCACTCTCTTGCCCAGGCTGAAGT 60

Qy        659 ACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGA 718
              ||||||  |   | |    | ||||||||  ||||   |||||||  | |||||    |
Db         61 GCAGTGGCGGGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCACTTCTTCTG 120

Qy        719 CCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAA 778
              | ||||||  ||   ||||||||||  |||  ||   |||||||  ||||||  || ||
Db        121 CTTCAGCCCCCTGAGTAGCTGGGATTACAGGCACCCACCACCAGGCCTGGATACTT--TT 178

Qy        779 AAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGC 838
              | |||||     | | ||||  ||  |||  ||||||||||  ||||  |||||||| |
Db        179 TTGTATTTTTAGTAAAAATGGGGTTTGCTATATTGGCCAGGCTGGTCTCAAACTCCTGAC 238

Qy        839 TTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCA 898
              |||  |  ||||  |||  |||   ||  | ||||||||||||||||||||||   ||| ||||||
Db        239 CTCAGGTGATCCACCTGCCTAAGCCTGCCAAAGTGCTGGGATTACAGGCATGAACCACCA 298

Qy        899 TGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAAC 958
              ||   | ||   |||  ||||||  | |          ||      | | |    |
Db        299 TGCCCAGCCTCTTTATTCTTTTTAATATTATCACAAGTTATTTATTTATTTATTTATTTA 358

Qy        959 AGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTC 1018
              |||  |    |      |||    ||   | |  ||     | | ||    | |
Db        359 TTTATTTATTTTTTGAGATGGAGTTTTGCCCTTGTTGCCCAAGCTGGAGTGCAATGGCAC 418

Qy       1019 AACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAA 1078
              | |     |||||||||||||||||  |||||||||||||||:|||||||||||||| |
Db        419 GATC-----TCAGCTCACTGCAACCTCCACCTCCCGGGTTCAAGTGATTCTCCTGCCTCA 473

Qy       1079 GCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTTGTATT 1138
              ||||||||||||||| ||||||  ||     ||| |||||  |||||||   |||||  |||
Db        474 GCCTCCTGAGTAGCTGGGATTACAGGCACATGCCATCATGCCTGGCTAA--TTTTGCATT 531

Qy       1139 TTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAG 1198
              ||||| |||||  | ||||  ||||||||  ||||||||||  ||||||||||||| |
Db        532 TTTAGTAGAGACAGAGTTTCTCCATGTTGGTCAGGCTGGTCTTGAACTCCTGACCTCAGG 591

Qy       1199 TGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTG 1258
              |||||||||  |||| | ||||||||||||||| |||||||| |||| ||| |||||
Db        592 TGATCTGCCCTCCTCGGCCTCCCAAAGTGCTGGGATTATAGGCATGAGCCACCGTGCCCA 651

Qy       1259 GCCGATTACTGTCTATTTTCTTTATTGCTATATC 1292
              |||  |  |      | ||  |  || || | ||
Db        652 GCCATTATCACAAGTTATTTATACTTCCTTTCTC 685


RESULT 13
CR613629
LOCUS       CR613629               3092 bp    mRNA    linear   HTC 21-JUL-2004
DEFINITION  full-length cDNA clone CS0DI039YN19 of Placenta Cot 25-normalized
            of Homo sapiens (human).
ACCESSION   CR613629
VERSION     CR613629.1  GI:50494436
KEYWORDS    HTC; CNSLT_cDNA.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 3092)
  AUTHORS   Li,W.B., Gruber,C., Jessee,J. and Polayes,D.
  TITLE     Full-length cDNA libraries and normalization
  JOURNAL   Unpublished
  REMARK    Contact : Feng Liang Email : fliang@lifetech.com URL :
            http://fulllength.invitrogen.com/ InVitroGen Corporation 1600
            Faraday Avenue
REFERENCE   2  (bases 1 to 3092)
  AUTHORS   Genoscope.
  TITLE     Direct Submission
  JOURNAL   Submitted (20-JUL-2004) Genoscope - Centre National de Sequencage :
            BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
            - Web : www.genoscope.cns.fr)
COMMENT     1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime
            end enriched, double-strand cDNA was digested with Not I and cloned
            into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library
            was normalized. Library was constructed by Life Technologies, a
            division of Invitrogen.
FEATURES             Location/Qualifiers
     source          1. .3092
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="CS0DI039YN19"
                     /tissue_type="Placenta Cot 25-normalized"
                     /plasmid="pCMVSPORT_6"
ORIGIN

  Query Match             11.6%;  Score 290.6;  DB 4;  Length 3092;
  Best Local Similarity   69.7%;  Pred. No. 3.7e-17;
  Matches  466;  Conservative    1;  Mismatches  190;  Indels   12;  Gaps    5;

Qy        594 ACAGTTTGGGTTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCA 653
              | | |||   |||||||||  || ||  ||| |||||||  ||||||||||||||||||||
Db       1750 AAATTTTTTTTTTTTTTTTTTTTTTTTTGAGACAGAGTCTTGCTCTGTCACCCAGGCT 1809

Qy        654 TGAGCACAGTGGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCT 713
              |||  ||||||||    | ||  | ||||||||  |||||  |||||||    | |||  |||
Db       1810 GGAGTGCAGTGGTG---CGATCGCTCACTGCAACCTCAGCCTCCTGGATTTAAGTGATTC 1866

Qy        714 GCTGACCTCAGCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAAT 773
              |    |||||||||||||||||||||||||||||  |||   | ||||||||||| ||| |||||||
Db       1867 TCCTGCCTCAGCCTCCCAAGTAGCTGGGATTACAGGTGTGCACCACCATGCCCGGCTAAT 1926

Qy        774 TAAAAAAATTTTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTC 833
              |       | ||||| || |||||  |||| | || | ||||||||||| ||||   |||||
Db       1927 TTTTTGTA-TTTTTAGTGGAGACAGGGTTTCACCATGTTGGCCAGGCTGGTCTCGAACTC 1985
```

```
Qy        834 CTGGCTTCAAGCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGC 893
              ||| | ||  |  |||| ||| |||  || ||||||||||||||||||||||| ||||||
Db       1986 CTGACCTC--GTGATCCGCCTGCCTCAGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGC 2043

Qy        894 CACCATGTGCGGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCAC 953
              ||||| |   ||| |         || |      | |    |         ||
Db       2044 CACCACGCCTGGCCCCAGTCCAAAATATTTAAAGATTGTTTCCTTAGTGTCTTGAAGTTT 2103

Qy        954 AGAACAGGGATTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATC 1013
              | |||      | | | | ||||     |    | |          |  |   ||  |
Db       2104 TGCACA-----AAATTCTTTTTTTTGAGATGGAGTCTCACTCTGTCACCCAGGCTGGAGT 2158

Qy       1014 ACCTCAACCTCCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTG 1073
              |      | |     |||||||||||||||||||||||| |||||||||: |||||||
Db       2159 GCAGTGGCGTGATCTTGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCAATTCTCCCA 2218

Qy       1074 CCTAAGCCTCCTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAATTTTTT 1133
              ||| |||||||  |||||||| |||||  |||| |||||| | ||| ||| |||||
Db       2219 CCTCAGCCTCCCAAGTAGCTGGGATTACAGACGTGTGCCACCATACCTGGGTAA-TTTTT 2277

Qy       1134 GTATTTTTAGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACC 1193
              | ||||||||  |||| || |||| ||||||||| ||||| ||||||  ||||||||||
Db       2278 GCATTTTTAGTGGAGAGGGAGTTTCACCATGTTGGCCAGGTTGGTCTTGAACTCCTGACC 2337

Qy       1194 TCAAGTGATCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGT 1253
              ||| ||||||  |||||||| | |||||||||||||| |||| |||| |||| || ||
Db       2338 TCAGGTGATCCTCCTGCCTCGGCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACCGT 2397

Qy       1254 GCCTGGCCG 1262
              ||  ||||
Db       2398 GCTCAGCCG 2406
```

```
RESULT 14
BQ681302/c
LOCUS       BQ681302                 855 bp    mRNA    linear   EST 15-JUL-2002
DEFINITION  AGENCOURT_8034339 NIH_MGC_112 Homo sapiens cDNA clone IMAGE:6213772
            5', mRNA sequence.
ACCESSION   BQ681302
VERSION     BQ681302.1  GI:21793981
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1  (bases 1 to 855)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: DCTD/DTP
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2380  row: a  column: 05
            High quality sequence stop: 649.
FEATURES             Location/Qualifiers
     source          1. .855
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6213772"
                     /tissue_type="melanotic melanoma, cell line"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_112"
                     /note="Organ: skin; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN
```

```
  Query Match           11.5%;  Score 288.2;  DB 5;  Length 855;
  Best Local Similarity   69.2%;  Pred. No. 1.3e-16;
  Matches  471;  Conservative    1;  Mismatches  161;  Indels   48;  Gaps    4;

Qy        606 TTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGTGG 665
              ||||||| || || |||  |||||||||||||| ||||||||||||| |  |||  ||||||
Db        855 TTTTTTTTTTTTTTTTTTTTAGACAGGGTCTTGATCTGTCACCCAGACTGGAGTGCAGTGG 796

Qy        666 TGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCAGC 725
              ||  | ||||| |||||| | | |||| ||| |  |||||  |  ||       || |||||
Db        795 TGAGATCATAGCTCACTGTAACNTCAAACTCTTCGGCTCAGGCAATACTTCCACTTCAGC 736

Qy        726 CTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATTTT 785
              |||||| |||||||||||||| || |    ||||||  ||  ||||| |     || |||
Db        735 TTCCCAAATAGCTGGGACTACAAGTGCATGCCACCATACCCAGCTAAAT-GTTTAAGTTT 677

Qy        786 TTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAAGC 845
              ||||||||||  ||||||| | | |||   ||||||  ||||||||||||||||||||
Db        676 TTTGTAGAGATGGGGTCTTGCCATGTTTCNCAGGCTGCTCTTAAACTCCTGGCTTCAAGT 617

Qy        846 AATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGCGG 905
              |||||||  ||||  || ||||||||||||||||| |||||  |||||  |||  |||   ||
Db        616 GATCCTCCCACCTCAGCCTCCCAAAGTGCTGGGATGACAGGTGTGAGACACAATGCCTGG 557

Qy        906 CTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGATT 965
              || ||   |   |  |  |   | ||||  | | |  ||
Db        556 CTTCTCTCTCTCTCTCTTTTTTTTTCTGAAACAGAGT---------------------- 518

Qy        966 ACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCTCC 1025
                          ||| || ||   | |   |   || | ||    ||| |
Db        517 -----------------TTTGTTCTTGTTGCCCAGGCTGCAGTGCAATGGCTCGA----- 480

Qy       1026 GTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTCCT 1085
              |||||||||||||||||||| ||||||||||||||:|||||||||||||| |||||||
Db        479 TTTCAGCTCACTGCAACCTCTACCTCCCGGGTTCAAGTGATTCTCCTGCCTCAGCCTCCC 420

Qy       1086 GAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA----TTTTTTGTATTTTT 1141
              ||||||||| |||||| |||    |||||| || ||||| |    ||||||||||||||
Db        419 AAGTAGCTGGGATTACAGGCGCCTGCCACCACGCCTGGCTGACTTTTTTTTTGTATTTTT 360

Qy       1142 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1201
              || |||||||||||||| |||||| || ||||||||   ||  ||||||||| |||| ||||
Db        359 AGTAGAGATGGGGTTTCACCATGGTGGCCAGGCTGACCTTGAACTCCTGATCTCAGGTGA 300

Qy       1202 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1261
              ||  ||  |||  | |||||||||||||| |||| ||| |||||  |||||||| ||||
Db        299 TCCACCCACCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACTGTGCCCGGCC 240

Qy       1262 GATTACTGTCTATTTTCTTTA 1282
                | ||  |  | ||  | |
Db        239 AGCTTCTTGCCTTCTTACTGA 219
```

```
RESULT 15
BQ710315/c
LOCUS       BQ710315                 769 bp    mRNA    linear   EST 16-JUL-2002
DEFINITION  AGENCOURT_8351732 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282212
            5', mRNA sequence.
ACCESSION   BQ710315
VERSION     BQ710315.1  GI:21849214
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
```

```
REFERENCE   1  (bases 1 to 769)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Dr. Mark Watson
             cDNA Library Preparation: Rubin Laboratory
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Agencourt Bioscience Corporation
             Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM2476  row: d  column: 21
            High quality sequence stop: 549.
FEATURES             Location/Qualifiers
     source          1. .769
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:6282212"
                     /lab_host="DH10B (phage-resistant)"
                     /clone_lib="NIH_MGC_113"
                     /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                     EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                     into EcoRI/XhoI sites using the following 5' adaptor:
                     GGCACGAG(G). Library constructed  by Ling Hong in the
                     laboratory of Gerald M. Rubin (University of  California,
                     Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                     Superscript II RT (Life Technologies). Note: this is a
                     NIH_MGC Library."
ORIGIN

  Query Match            11.5%;  Score 287.6;  DB 5;  Length 769;
  Best Local Similarity  67.0%;  Pred. No. 1.6e-16;
  Matches  453;  Conservative    1;  Mismatches  215;  Indels    7;  Gaps    3;

Qy        604 TTTTTTTTTGTTGTTGTTTAGAGACAGGGTCTTGCTCTGTCACCCAGGCATGAGCACAGT 663
              ||||||||| || || | | ||||||| ||||  ||||||| |||||||  |||  ||||
Db        768 TTTTTTTTTTTTTTTTTATTGAGACAGAGTCTCACTCTGTCCCCCAGGCTGGAGTGCAGT 709

Qy        664 GGTGCAACCATAGGTCACTGCAGCCTCAACCTCCTGAGCTCAAGGGATCTGCTGACCTCA 723
              || |  | |   | |||||||| ||||  ||||||| | |||||  |||    |   |||||
Db        708 GGGGTGATCTCGGCTCACTGCAACCTCTGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCA 649

Qy        724 GCCTCCCAAGTAGCTGGGACTACGAGCGTGCACCACCACGCCTGGCTAATTAAAAAAATT 783
              ||||||| ||||||||||||||    |||  | ||||   |||  |    | ||  | |   |
Db        648 GCCTCCCGAGTAGCTGGGACTATAGGCGCCCGCCACTGTGCCGGCTAATTTTTATA---T 592

Qy        784 TTTTTGTAGAGACTGGGTCTTACTACGTTGGCCAGGCTTGTCTTAAACTCCTGGCTTCAA 843
              |||| |||||||| |||| |   | | |||||||||||||| ||||  || ||||  | |||
Db        591 TTTTAGTAGAGACGGGGTTTCCCCATGTTGGCCAGGCTGGTCTCGAATTCCTAACCTCAG 532

Qy        844 GCAATCCTCCTACCTTGGCATCCCAAAGTGCTGGGATTACAGGGGTGAGCCACCATGTGC 903
              |  |||| |||  ||  || ||||||  |||||||||||||| ||||||||||||  |
Db        531 GTGATCCGCCTGGCTCAGCCTCCCAACATGCTGGGATTACAGGTGTGAGCCACCATGCCC 472

Qy        904 GGCTACTTATTTCTTTACATTCCATCTTTCCAATAGAATGTAAGATCCACAGAACAGGGA 963
              ||||   |   ||||   |    |||  |   |          | | ||     |
Db        471 GGCTGTGTTACACTTTTTGTGGTGTCTCACAGGTCTCTGAGGCTCTGCTCATTTTTGTTT 412

Qy        964 TTACTGCCTATTTTCTTCCTTTCTTTTTTGAGACAGAGTCTCACTTCATCACCTCAACCT 1023
              ||  |     |||| |          | |                |||  |   |
Db        411 TTGTTTTTATTTTTTTGAGACAGAGTCTCTTTCTGTCACTAGACTGGAGTGCAGTGGTGC 352

Qy       1024 CCGTTCAGCTCACTGCAACCTCTGCCTCCCGGGTTCAAGYGATTCTCCTGCCTAAGCCTC 1083
                || ||||||||||||||||||||| | |||||||:||||||| ||||| ||||||
Db        351 AATCTCGGCTCACTGCAACCTCTGCCTCCAGAGTTCAAGCGATTCTCTTGCCTCAGCCTC 292

Qy       1084 CTGAGTAGCTGGAATTACAAGCGTGCACCACCATGCTTGGCTAA--TTTTTTGTATTTTT 1141
              |  |||||||| |||||| ||     |||||||||    ||||  ||||||||||||||
Db        291 CCAAGTAGCTGGGATTACAGGCACATGCCACCATGCCCAACTAATTTTTTTTGTATTTTT 232

Qy       1142 AGCAGAGATGGGGTTTTACCATGTTGCCCAGGCTGGTCTCAAACTCCTGACCTCAAGTGA 1201
              || |||||||||||||| |||||||||| ||||| ||||||   | ||| |||||||  ||||
Db        231 AGTAGAGATGGGGTTTGACCATGTTGGCCAGGATGGTCTAGATCTCTTGACCTC--GTGA 174

Qy       1202 TCTGCCTGCCTCAGTCTCCCAAAGTGCTGGAATTATAGGCGTGAGTCACTGTGCCTGGCC 1261
              ||   |  |||  | |||||||||||||  |||| |||||||| ||||| ||||||||
Db        173 TCCAACCACCTTGGCCTCCCAAAGTGCTGAGATTACAGGCGTGAGCCACTGCGCCTGGCC 114

Qy       1262 GATTACTGTCTATTTT 1277
              | | ||  | || |
Db        113 GGCTCCTCCTTCTTAT 98


Search completed: February 17, 2006, 16:36:09
Job time : 9879 secs
```